CROSS-MODAL CONTACTS BETWEEN GRAPHEMES AND PHONEMES

EXPLORATIONS IN BIMODAL PROCESSING

TON DIJKSTRA

CROSS-MODAL CONTACTS BETWEEN GRAPHEMES AND PHONEMES

EXPLORATIONS IN BIMODAL PROCESSING

CROSS-MODAL CONTACTS BETWEEN GRAPHEMES AND PHONEMES

EXPLORATIONS IN BIMODAL PROCESSING

een wetenschappelijke proeve

op het gebied van de sociale wetenschappen,

in het bijzonder de psychologie

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit te Nijmegen,
volgens besluit van het college van decanen
in het openbaar te verdedigen
op maandag 26 maart 1990,
des namiddags te 3.30 uur

door

Antonius Franciscus Jacobus Dijkstra
geboren te Tilburg

Promotores: Prof. dr. R. Schreuder
Prof. dr. G.A.M. Kempen

Co-promotor: Dr. U.H. Frauenfelder
(Max-Planck-Institut für Psycholinguistik)

NICI Technical Report 90-02
ISBN 90-9003316-5/CIP/NUGI 941

Contents

Acknowledgements

Deprived of light and communication the experimental psychologist spends long hours sitting silently, listening to the clicks of growing data. But, of course, a larger part of scientific research is a social affair: Ideas are shaped and sharpened in the interaction with colleagues and experimental subjects.

This thesis has, first of all, benefited much from Never Ending discussions with Rob Schreuder and Uli Frauenfelder. To Gerard Kempen I am especially grateful for providing me with the warm comfort of sufficient "infrastructural support": enough time and freedom to finish a thesis on a theme that was, to paraphrase a Dutch expression, a strange duck in a language-technological pond.

I am further grateful to Dirk Vorberg, for teaching me the basics of computer-modeling and showing me how, to my excitement and enthousiasm, mathematics can be fruitfully used in psycholinguistics. Without his chalk and talk, part of this thesis would never have been written. Thanks are also due to Ar Thomassen and Pienie Zwitserlood, who gave detailed comments on the text. I would like to thank several other people who helped me in various ways: Agnes Bolwiender and Hans Oskamp (for their assistance in running and analyzing the experiments); Sylvia Aal (for her substantial aid and advice on the lay-out of the text); Henk van Jaarsveld, Herbert Schriefers, Ardi Roelofs, and Gerard van Galen (for extensive discussions); Jaap Havinga, Ad van de Ven, Gerard van Breukelen, and Clemens de Leeuw (for mathematical and FORTRAN-insights). I am further indebted to all members of the MPI-Stipend Group; and, of course, to my experimental subjects and friends (two overlapping groups, certainly at the beginning of my research...).

Two institutions must not remain unmentioned in this word of thanks: the Max-Planck-Institute for Psycholinguistics, for offering me a stipend and laboratory facilities; and the University of Nijmegen, for making it possible to continue both my thesis research and my teaching.

Last, but certainly not least, I want to express my gratitude for lots of back ups and unceasing moral support ... "aan Haar Zonder Wier ...".

1

Introduction

One of the basic tenets of psycholinguistics is that the processing of language involves a series of computations upon mental representations of the incoming sensory information (Fodor, 1975, 1981). Indeed, the psycholinguistic enterprise can be described as an attempt to determine the temporal and structural characteristics of the various types of computations and representations that figure in linguistic processing.

Most researchers assume that linguistic processing is performed by a complex hierarchical system that is especially and only equipped to handle linguistic information. In Fodor's influential conception, language is considered to be a *cognitive module*, i.e., a "domain specific, innately specified, hardwired, autonomous and not assembled" computational system (1983, p. 37). Language is seen as a "vertical faculty", a distinct processing system that, while it transforms linguistic input into some central internal representation, does not interact or interfere much with other cognitive processing systems (like those for mathematical reasoning or listening to music), since it does not compete with them for "horizontal" resources such as memory or attention.

One of the defining properties of a module is that it is domain specific and accepts only a restricted type of input, e.g., during the perceptual analysis of speech not all acoustic signals are operated upon, but only those that are taken to be utterances (Fodor, 1983, p. 49). In case of language the interesting problem arises what should be understood by "domain specificity", since quite diverse sorts of input are acceptable to the language processing system, not only different in form (such as handwriting and print) but also in modality (such as print, speech and gestures). Fodor is inclined to treat the whole of language as the domain of one module (Marshall, 1984, p. 221, 223), and within this module he assumes cross-modal linkages to handle, e.g., the processing of both visually and auditorily presented linguistic information (Fodor, 1983, p. 132).

However, according to an alternative viewpoint, the processing of visual and auditory linguistic stimuli – to which I limit myself here – is carried out by systems that can to a large extent be considered modules on their own. This alternative view argues that several properties of modules do not seem appropriate for language as a whole (Shallice, 1984). For example, while speech is thought to be processed by an innate system that evolved during human evolution, print clearly is not (Liberman, 1988). However, since the reading system may well be computationally autonomous and partly mandatory, it can be considered a processing system that is separable from others.

Another argument for assuming some autonomy for the visual and auditory language processing systems is found in the vast differences in the characteristics of the input signals and channels of the two systems (cf. Nickerson, 1981). Visual and auditory linguistic signals differ in both temporal and structural aspects. Auditory information arrives bit by bit, slowly accrues over time, and is evanescent. The speed of auditory information arrival is determined by the speaker. Visual information, however, is present in chunks, appearing "in a wink of the eye", continuously available for analysis or re-analysis, while the speed of information processing is in principle under control of the reader.

Structural characteristics are also remarkably different. In our alphabetic writing system, printed texts usually consist of clearly recognizable words, built up with discrete elements of high contrast (letters). The auditory speech signal, however, varies over time in strength, clarity and content: E.g., information about a particular phoneme is not available in an "all-or-none" fashion, but it grows and decays, influenced by the context in which it appears (coarticulation phenomena, sandhi). Some researchers have therefore likened the speech signal to a code, from which the message can only be deciphered in an indirect and complex way (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

Despite such differences in input signals, these visual and auditory processing systems are both meant to recognize forms of *language*, and it seems undisputed that in the end they give rise to a common semantic representation. To decide to which extent the visual and auditory processing systems should be seen as submodules within language as a whole, or as different modules on their own, we need to learn more about specific aspects of their relationship. Two important issues are the following.

1. How similar or different are these subsystems we categorize under the heading "language"? Do they have a comparable architecture? How do temporal and structural differences in input signals and channels influence linguistic processing?

2. Do the subsystems, though they start out with quite different types of input, make contact or even converge at more abstract levels? Do they share the same knowledge sources (e.g., of lexical, syntactic and semantic knowledge)? Do they share "horizontal" resources such as memory or attention?

1.1 Investigating visual-auditory similarities: two approaches

The similarities between the visual and auditory linguistic systems have been investigated by two different approaches. First, psycholinguists have compared experimental data and theoretical views that are specific to each modality; and, second, they have run experiments in which linguistic processing in one modality is made to resemble that in the other modality through clever manipulation of the stimulus presentation.

A comparison of the theory and data available for the visual and auditory modality separately may bring out similarities of the two processing systems. Examples may be found in analogous effects on a word's recognition process by its orthographic or phonological "neighbors" (e.g., Frauenfelder, 1990), in similarities of visual or auditory access codes (Taft, 1986), or in similar interactions of word recognition and sentence context (Norris, 1986). However, this type of comparison is limited. The main problem is to define what is comparable or should be compared in the two modalities. Experimental results are to some extent task dependent, and tasks are not directly comparable across modalities (take, e.g., visual and auditory lexical decision). Stimulus characteristics are not directly comparable either (e.g., parts of the word TONGUE auditorily come in gradually over time, but visually probably in only one chunk).

Other research has taken the second approach, examining whether linguistic processing in one modality can be simulated by manipulating characteristics of processing in the other modality (e.g., Blosfelds, 1981). An important characteristic of the auditory modality is that speech information comes in over time and is evanescent. This characteristic may be mimicked in visual presentation by presenting words or letters on a computer screen for only a limited period of time, one after the other. A characteristic property of visual word recognition already mentioned is that several letter units constituting a word appear more or less simultaneously. This feature of visual processing may be imitated by presenting time-compressed speech (more speech in less time) while keeping other characteristics of speech constant.

Though some interesting results have been obtained in this way, it is hard to understand how similarities in experimental results obtained after such manipulations must be interpreted. Do the results indeed indicate that the two processing systems function analogously, and that the different behavior that we normally observe is caused only by differences in the input? Or do the manipulations change the normal course of processing by tinkering with inherent characteristics of visual or auditory processing? In the light of the extreme flexibility of the human language processing system it seems dangerous to interpret these types of experiments without convergent evidence from other tasks.

1.2 Investigating visual-auditory contacts: two approaches

The approaches just discussed in a sense compare "vertical" properties of different language subsystems. Indeed, most of the models that describe linguistic processing are "vertically oriented" in that they follow the flow of linguistic information from bottom (signal) to top (some abstract mental representation). The *relation* between the various subsystems has seldom been an explicit focus of research. Yet, knowledge of contacts and convergences of subsystems is indispensable for a complete picture of the internal structure and the temporal aspects of the linguistic processing system.

This thesis could be considered as part of a project to investigate in more detail the intricate relationship between the visual and auditory processing systems. Because of the generally acknowledged hierarchical nature of these systems, it seemed advisable to start such a project by examining contacts between small representational units, which may function as building blocks for more complex representations. In this thesis I consider contacts between grapheme and phoneme representations, as they are the smallest visual and auditory representational units which stand in a meaningful relation. Furthermore, this domain is one of the few where an exploration of cross-modal contacts has already started.

I will often refer to grapheme and phoneme representations with the term "sublexical". The term is used in this thesis to indicate any representation smaller than the word. Such a representation may be called "prelexical" if it is computed as an intermediary representation before word recognition takes place. Thus, in this thesis the term "lexical" is used to refer to word representations (e.g., "lexically mediated" is taken to mean "mediated by word representations").

If we assume that various sublexical representational units become activated during linguistic processing, *contact* can be defined as activation of representations in one modality by those in another (e.g., grapheme by phoneme representations). For the special case in which a representation is activated that is common to the two modalities, I would like to reserve the term *convergence*.

A number of specific questions concerning contacts between visual and auditory sublexical representations can now be posed:

1. Does any cross-modal activation spreading occur between grapheme and phoneme representations during the processing of visual and/or auditory linguistic material?

2. Is any occurring cross-modal activation bidirectional, i.e., does it spread both from the visual to the auditory domain and vice versa?

3. What can be said about the temporal aspects of any existing cross-modal activation?

4. Can cross-modal activation between graphemes and phonemes (if it occurs) be only facilitatory, or inhibitory as well?

5. Does cross-modal activation (if it occurs) take place automatically, i.e. fast and without conscious control of the subject (Posner & Snyder, 1975a, b)?

An answer to these questions will not only enhance our knowledge of how the visual and auditory processing systems relate at a sublexical level, but can also be seen as a step towards the development of a generalized model of word recognition, incorporating both the visual and the auditory subsystems.

Two types of approach are fit to look for contacts between linguistic processing in different modalities. Experiments involving only visual stimuli have examined effects of phonological recoding in word recognition and sentence processing (e.g., Tanenhaus, Flanigan, & Seidenberg, 1980; Jakimik, Cole, & Rudnicky, 1985; Treiman, Freyd, & Baron 1983; Black, Coltheart & Byng, 1987). Bimodal priming experiments have investigated cross-modal activation of lexical or semantic representations (e.g., Marslen-Wilson, 1990; Zwitserlood, 1989).

Within the first approach, visual lexical decision experiments have shown that the time needed to respond "no" to nonwords that are homophonic with real words ("pseudohomophones", such as BRANE) is longer than to other, nonhomophonic, nonwords (such as BRAME) (e.g., Coltheart, Davelaar, Jonasson, & Besner, 1977). This result has been interpreted as evidence that phonological information becomes active during visual word recognition. A similar interpretation has been given to the finding that in a categorization task subjects take more time to reject the word PAIR, homophonic to PEAR, as a member of the category FRUIT than other homophonic nonexemplars, such as TAIL (Meyer & Gutschera, 1975). However, the results of experiments like these have stirred up discussion (e.g., Humphreys & Evett, 1985), because they often failed to disentangle effects of orthographic and phonological similarity, and had trouble to distinguish phonological effects arising before, after or even without lexical information becomes available (Patterson & Coltheart, 1987). Recent research with more adequate control conditions seems to provide some evidence that phonological information may indeed become automatically available before visual word recognition (e.g., Van Orden, 1987). However, even in this research it has not been demonstrated that visual and auditory processing activate one and the same type of phonological representations.

A more promising way to address cross-modal issues lies in the application of tasks involving bimodal stimulus presentation. Such tasks have used both sequential and simultaneous presentation of visual and auditory stimulus material. An example of the sequential approach is repetition priming, a technique in which a target stimulus (e.g., a word) is presented first in one modality, and later in the other modality. Effects of the prime on some performance measure regarding the target can be interpreted in terms of contacts or convergence (e.g., Jackson & Morton, 1984; Monsell, 1985). However, a general problem of this priming technique is that it is concerned with the products of processing, not processing itself. It is therefore not clear where in the processing system repetition priming effects arise. Forster and Davis (1984) hold that part of the obtained priming effects must be ascribed to episodic memory factors.

When stimuli in the two modalities are presented closer together in time, this problem disappears. We may distinguish situations where subjects must keep track of one of two simultaneous messages (focused attention), and where they keep track of each of two simultaneously presented messages or signals (divided attention). In the *focused attention* task, the subject detects a target in one modality, while an accessory stimulus appears in the second modality just before, simultaneously with, or shortly after the target. By manipulating the relationship between the context stimulus and the target (e.g., presenting name-identical words or not), cross-modal activation effects can be examined. It may be assumed that the general, non-specific effects of visual accessories will not differ very much when one stimulus is presented rather than another (e.g., letter A will have the same non-specific effect as letter P).

In situations requiring *divided attention*, the task is to attend to several simultaneous target channels at once, responding to each as needed. The time to detect a target in one modality may be differentially influenced by a target or non-target stimulus in the other modality. Manipulation of the onset asynchrony of the visual and auditory stimuli can provide information on the risetime or directionality of activation (visual to auditory or auditory to visual). How the subjects divide their attention over the two modalities can be investigated by comparing reaction times (RTs) to single-channel conditions with those to bimodal conditions in which a target is combined with a neutral stimulus.

Within the bimodal approach, the structural and temporal relationship between prime and target can easily be manipulated, and the reaction to a target stimulus gives a rather direct reflection of cross-modal effects exerted by a prime. Because of these advantages and the disadvantages of the other approaches reviewed, the experiments reported in this thesis were all bimodal, either of the focused attention variant or of the divided attention variant.

1.3 Structure of the thesis

As a basis for the research to be reported, Chapter 2 of this thesis gives an overview of the literature on bimodal experiments that concern the issue of sublexical contacts between the visual and auditory processing systems. Evidence for cross-modal contacts between graphemes and phonemes is presented, but it is concluded that most of the questions posed above remain unanswered.

Chapter 3 presents three focused attention experiments that address some of these questions. In particular, the experiments attempt to demonstrate grapheme-to-phoneme activation, and examine its time course and automaticity. In these *auditory vowel-detection* experiments, Dutch subjects make a forced choice on the identity of the vowel (e.g., /a:/ or /e:/) in an auditorily presented syllable (e.g., /pe:/, /a:/, or /e:k/). Visual letter primes (e.g., P, A, or E) are presented before, during or after presentation of the syllable. In the *indirect priming* manipulation, the relationship between the letter and the consonant adjacent to the vowel is varied, resulting in consonant-congruent conditions (e.g., letter P combined with auditory syllable /pa:/) and consonant-incongruent conditions (e.g., P with /ka:/). In the *direct priming* manipulation, the relation between the letter and the target vowel itself is varied, leading to vowel-congruent conditions (e.g., A with /a:k/) and vowel-incongruent conditions (e.g., A with /e:/). After the results have been put in a theoretical framework, they are compared with those obtained in Stroop-experiments and their consequences for word recognition models are discussed.

Chapter 4 investigates the issue of automatic bidirectional activation of graphemes and phonemes in a go/no-go divided attention task. In three *bimodal vowel-detection* experiments, Dutch subjects detect visual and auditory target vowels.[1] Some bimodal conditions are *redundant*, i.e., they contain two targets, which are either congruent (e.g., visual A, auditory /a:/) or incongruent (e.g., U, /a:/). The visual and auditory stimuli of the congruent redundant conditions are name-identical in Dutch. Temporal aspects of cross-modal activation are examined by varying the stimulus onset asynchrony (SOA) of visual and auditory component stimuli. Results for bimodal non-redundant conditions (that include one target and one neutral stimulus) provide information about general effects of visual and auditory stimuli, which may influence the bimodal redundant conditions as well.

It is assumed here that in a redundant trial the visual and auditory targets are involved in a race for identification, and that the winner of the race evokes the response. Furthermore, if one target is presented earlier in time than the other, it will more often win the race. Thus, if a visual target precedes an auditory target, it will be reacted to more often than when both targets are presented simultaneously. With respect to a situation in which the auditory target leads, the difference in the number of reactions to the visual target will be even larger. The race between the two signals may be modified by cross-modal activation effects occurring at representation, decision or response levels. A statistical method developed by J. Miller (1982) can often provide evidence that such "co-activation" effects are present in the data. If grapheme and phoneme representations activate each other, coactivation effects should be present in the congruent conditions, but not necessarily in the incongruent conditions.

More direct evidence for cross-modal activation effects at the representational level is obtained in the following way. The RTs in the redundant conditions are corrected for differences in processing time among the different letter and speech sound target stimuli by a method that uses the obtained single-channel RTs. After the RTs have been adapted, the congruent conditions should show more facilitation than the incongruent conditions. The SOA-manipulation here offers an indication whether the effects are bidirectional. Under the assumption that subjects most often react to the stimulus that is first presented, cross-modal effects will reflect phoneme-to-grapheme activation if the visual target precedes the auditory, and grapheme-to-phoneme activation if the auditory target precedes the visual.

However, it could be argued that this SOA-manipulation does not provide conclusive evidence in favor of bidirectional cross-modal activation effects. Since it is not known to which targets subjects react to in the go/no-go task, it cannot be excluded that they develop a general strategy favoring reactions to the visual modality (visual dominance effect) instead of reacting more often to the first presented target. Chapter 5 examines this possibility and at the same time tries to obtain more empirical support for bidirectional activation effects in a new divided attention paradigm, that of *modality decision*. In a two-choice situation, subjects determine the modality in which they first detect a target. Though more research is required to evaluate the merits and pitfalls of this paradigm, the results are generally consistent with those obtained by the two other types of tasks.

In Chapter 6 the results of the three task types (bimodal and auditory vowel-detection, and modality-decision) are compared and integrated into a more complete view on structural and temporal aspects of grapheme-phoneme contacts in sublexical bimodal processing. The questions raised in this Introduction are reviewed and, as far as possible, answered. Bimodal redundant conditions in the three types of experiment are compared in terms of the size and consistency of their results. Bimodal neutral and single-channel conditions help an analysis of performance differences among the three task types (e.g., due to differences in attention allocation). The chapter relates the reported research to the domain of word recognition and concludes with suggestions for further research.

Finally, in Chapter 7 the foundation is laid for a computer model that simulates the empirical results reported in this thesis. Step by step the model is built up, incorporating views on processing reported in the literature (Sanders, 1983) as well as the constraints offered by the data. It is shown that an implemented version of the model can account reasonably well for the results obtained with the bimodal and auditory vowel-detection tasks. Furthermore, some information is given about the choice of parameters, the effect of variation in those parameters and the goodness of fit. This first exploration holds some promise for further developments of the model.

2

Research on bimodal sublexical processing

Few psycholinguistic studies have investigated the relationship between the auditory and visual processing systems, and little more than a handful of them examined cross-modal contacts of sublexical representations. This chapter critically reviews these studies, which involved either the divided or the focused attention paradigm. A well-known example of the divided attention approach were the "same-different matching" experiments performed in the sixties and early seventies by Posner (e.g., 1978) and colleagues. Though the review will focus on bimodal matching studies, all-visual matching experiments will briefly be considered because they were thought to support the hypothesis that a phonetic name code is retrieved during letter processing. Focused attention studies involved various other experimental techniques, like naming visual stimuli that were accompanied by auditory material. The chapter ends with a summary of what can be learned from the studies reviewed.

2.1 Results of divided attention tasks: same/different matching

Posner (1969) proposed that during the processing of a visual letter stimulus three types of representations or *codes* ("the format in which information is represented") are derived: a physical code, a name code and a semantic code. The physical code is modality-specific and different for the visual and auditory domain. The name code is common to both modalities and is called "phonetic" by Posner, a term adopted "to stand for the code that underlies the internal naming of visual and auditory stimuli" (p. 30). However, going beyond this neutral definition, Posner suggests that the "phonetic code" is intrinsically quite similar to the auditory (physical) code (p. 42). This is also seen in the following figure (from Posner, 1978, p. 31), in which a notation is used for the name code that is commonly applied to phonemes.[1]

Stimulus	visual		auditory		visual	
Physical Code	A	a	ã	b̃	B	b
Phonetic Name Code		/a/			/b/	
Semantic Code		vowel			consonant	

Figure 2.1. Three types of codes distinguished by Posner (1969).

As the figure shows, different physical codes are assumed to exist for upper and lower case letters and auditory stimuli, but common phonetic name codes and semantic codes for all three. Since this theoretical distinction between codes was initially derived from matching experiments involving only visual stimuli, I first consider these. It will be argued that, though the name code may indeed be phonetic, these all-visual experiments do not provide strong support for this view (as Posner claimed). Subsequently, I discuss evidence from bimodal matching experiments in favor of cross-modal contacts between sublexical representations.

In the same/different matching paradigm, subjects were presented with pairs of lower and/or upper case letters that belonged to the same or different categories (e.g., B, b or a), and the subjects' latency to decide whether the stimuli of a pair were similar or different with respect to a certain dimension was recorded. Subjects were to respond "same" in some conditions if the letters were physically identical (physical instruction), in other conditions if the letters had the same name (name instruction), or if they were members of the same semantic category (rule instruction).

Using the rule instruction, the matching time for physically identical letter pairs ("AA") was about 70-100 ms shorter than that for pairs having only the letter name in common ("Aa"). "It seemed reasonable to postulate that the retrieval of the phonetic information involved in the name was the source of this extra reaction time" (Posner, 1978, p. 31). This conclusion was based on two observations. First, the reaction "different" to "AB" was about 80 ms faster for physical instructions than for name instructions, indicating that the subjects' matching process varied with the type of instruction received. Second, under physical instructions reactions to "Aa" were no slower than to "Ac", while those to "EF" (physically similar) were slowed down. However, in my opinion it does not follow that the name code in the visual matching experiments should be "phonetic", since a case-independent graphemic code (cf. Evett & Humphreys, 1981) could account for the RT-pattern just as well.

In later experiments that employed a matching task with letter stimuli, it was consistently found that even when the subjects are matching on the basis of "name same", the match is faster if the stimuli are also physically identical. This name-physical disparity, as Proctor (1981) has termed it, has been interpreted as due to differences in latencies for making matches on the basis of a physical code as opposed to the name code. Proctor has further implicated name codes as the basis for the disparity in latency of same and different judgments. Typically, "same" judgments are made more rapidly than judgments of "different". In Proctor's general theory of matching behavior, this same-different disparity is attributed to competition and resulting inhibition between the two name codes that are activated when the stimulus pair is different as opposed to the single name code that is activated when the stimuli are identical or nominally the same.

2.1.1 Simultaneous matching

Studies with simultaneous bimodal matching provide further evidence for the existence of a physical and a name code, but most of them do not seem to require postulation of a name code that is phonetic in nature.

Posner and Mitchell (1967) report on a series of seven matching experiments, two of which were bimodal (cf. also Posner, Boies, Eichelman, & Taylor, 1969). In the bimodal experiments, responses were required to pairs of visual digits (0-9) presented one to each eye, pairs of auditory digits presented one to each ear, and a pair of audio-visual digits. Due to technical constraints, the onset of the visual stimuli was simultaneous, but for the visual-auditory pairs the auditory stimulus always led by 20 ms, and for the auditory-auditory pairs there was an asynchrony varying between 0 and 40 ms. Two experiments with (just) five subjects were run with different instructions, involving (only) 5 times 30 trials (15 "same" and 15 "different") for each of the three combinations. One experiment required the subject to respond "same" if the digits were both odd or both even. The results showed no RT-differences between auditory-visual pairs and same-modality pairs (though it seems from the figures in the article that auditory-auditory pairs initially evoked slower responses). The other experiment required subjects to respond "same" if the two stimuli were the same digit, and if not they were to respond "different". In this experiment auditory-visual matches took about 50 ms longer than visual-visual matches and 80 ms longer than auditory-auditory matches.

According to Posner, the combined results of these experiments support the hypothesis that the extraction of a common phonetic code was necessary in order to match a simultaneous visual and auditory digit in the last experiment. "The study suggests that obtaining the common phonetic code takes about the same time regardless of the modality from which it is extracted. It is particularly striking that varying input modality should require subjects to make the match based upon a phonetic code, since the learned correspondence of a visual digit and its auditory equivalent is so strong" (Posner, 1978, p. 42).

Actually, it seems to me that one can hardly expect subjects to match visual and auditory stimuli physically (as Posner and Mitchell themselves state: "Of course audio-visual pairs can only be equated at Node 2", i.e. that of the name code). Furthermore, the only conclusion that these results really seem to allow is that a common code was necessary in order to perform the task, but this code may just as well have been a visual code or some hypothesized abstract amodal code.[2] An implicit extra assumption made here is that the same name code is involved in the matching of digits and of digit names (number words). Finally, on the basis of Posner and Mitchell's experiments one cannot draw any conclusions concerning the automaticity or speed with which the common code became available. These problems seem to be directly tied to characteristics of the matching task: In it, the subjects are forced to transform both stimuli of a pair into a common format, but which type of format they choose or when exactly the match takes place cannot be determined on the basis of their responses. It may be that the type of code chosen for the match depends on characteristics of the subjects and the task situation. I will return to this issue when I discuss the results of some matching experiments by Wood (1974, 1977).

2.1.2 Successive matching

Involvement of a phonetic sublexical code is also claimed on the basis of studies in which there was a certain time interval (inter-stimulus interval or ISI) between the two stimuli of a matching pair. I will review only the most convincing studies that employed this successive matching technique.

Thorson, Hochhaus, and Stanners (1976) presented a single upper case letter for half a second, followed next by a blank field lasting from 0 to 2 seconds, and finally by a (non)matching letter. This second letter was also upper case, so in theory it would be possible for the subjects to make their matches based either on the physical or on the presumed phonetic code. The critical data were from the "different" responses that involved letter pairs that could be either visually or auditorily similar. RT-patterns showed no effects of acoustic confusability at very brief intervals, but from 1 second on strong interference effects started to arise. Visual similarity, however, had a great effect at short intervals, but by 2 seconds visually similar pairs were handled much faster than others. Similar results arguing for the influence of auditory similarity on visual matching were obtained by Dainoff (1970), and Dainoff and Haber (1970). These researchers found that name matches were affected by auditory similarity so that RTs were longer for similar pairs than for dissimilar pairs.

Cohen (1969) showed that for a longer ISI of 5 seconds RTs were only lengthened for items that were both visually and acoustically confusable, leading him to propose that comparisons are normally made in both channels. Posner (1978) suggests that for short ISIs there might be a race between codes, while at longer ISIs neither code is very active. However, even though these experiments are suggestive of the rise or fall of certain mental codes over (relatively long stretches of) time, they strictly spoken do not prove that the phonetic code involved in matching is *identical* to a code used in auditory perception.

A study by Boies (1971, reported in Posner, 1978) provides stronger evidence with respect to this point. Boies let subjects perform a letter matching task with intervals of either 0 or 2 seconds. On half of the occasions the letters matched physically, on the other half they matched only in name. For half the trials an irrelevant set of auditory "letters" (i.e. speech sounds) was also presented that the subjects were to hold in store during the trial. At the 2-second interval there was no influence of the irrelevant auditory set for physical matches, but for the phonetic matches there was substantial interference. These results are in line with the suggestion that two relatively independent codes exist following the representation of the visual letter. The relative advantage of the physical code over the phonetic code tends to decline over time, even though this does not mean that code is completely lost. Also, Posner interprets the study as an indication that the phonetic representation abstracted from a visual letter occupies the same system as the code extracted from speech sounds. If this were not so, the existence of the auditory memory load would not be expected to interfere with the phonetic code more than it interferes with the physical code.[3]

2.1.3 Studies criticizing Posner's code model

Many researchers have accepted Posner's conclusion that the studies just reported show the involvement of acoustic codes in letter matching, and they have tried to incorporate the experimental results in their models (e.g., Hardzinski & Pachella, 1980; Proctor, 1981; Proctor & Rao, 1983a, 1983b). However, from the time the experiments were run up till now, inconsistent results have been reported as well. According to Dainoff (1970) several all-visual matching experiments failed to obtain auditory confusability effects (Chase & Calfee, 1969; Cohen, 1968; Dainoff & Haber, 1967; Glucksberg, Fisher, & Monty, 1967; Kaplan, Yonas, & Shurcliff, 1966).

Boles (1981), and Boles and Eveland (1983) have presented evidence indicating that name codes are not typically involved in same-different matches, even when the matches are at the nominal level. Boles (1981) investigated the claim made by researchers that physical and name matches show visual field superiority effects in opposite directions, reflecting the operation of right-hemisphere visuospatial and left-hemisphere verbal processes. In seven experiments Boles failed to find any consistent lateralization effects. In another series of experiments, Boles and Eveland (1983) did not obtain evidence that name codes were involved in matching tasks and even could not replicate two experiments (Dainoff & Haber, 1970; Thorson, Hochhaus, & Stanners, 1976) that were supposed to have shown the involvement of name codes in same-different matches. Instead of name codes, Boles and Eveland found evidence for generation of visual or physical codes even when the to-be-matched stimuli were presented successively with relatively long (2 seconds) intervening intervals (also see Yeh & Eriksen, 1984).

Most detrimental to Posner's view that acoustic codes play a role in visual letter-matching are the results of two recent experiments by Carrasco, Kinchla, and Figueroa (1988). These researchers ran replications of the simultaneous matching study by Posner and Mitchell (1967), and the successive matching study by Thorson, Hochhaus, and Stanners (1976) (both reported above), using a much larger number of subjects. The original objective of the researchers was to investigate possible matching differences between English and Spanish subjects, expected because Spanish, unlike English, has a virtually invariant grapheme-phoneme correspondence. However, the two groups of subjects turned out to behave similarly, but very different from the subjects in the studies reported above. No evidence was found at all in favor of the hypothesis that acoustic codes were used in visual matching.

Not only all-visual matching experiments, but also bimodal matching experiments must be critically examined. Wood (1974, 1977) presents results from bimodal matching experiments that seem at least partially inconsistent with Posner's (1978) account. Letter pairs were constructed which were either high in auditory similarity and low in visual similarity (HALV), such as "ZC" (pronounced "zee - see"), or low in auditory similarity and high in visual similarity (LAHV), such as "OC" ("oh - see"). The expectation was that when stimuli were matched on the auditory dimension, HALV would lead to longer RTs than LAHV, but when matches would be made on the visual dimension, the opposite would obtain. The stimuli were presented visually in one block and auditorily in another. The visual stimuli and their auditory equivalents were presented for approximately 360 ms, with a stimulus onset asynchrony (SOA) of 1 second (from onset first stimulus to onset second stimulus), and an inter-trial interval (ITI) of 3 seconds. The following results were obtained for the different responses:

Similarity condition:	HALV	LAHV
Visual presentation	406	435
Auditory presentation	445	376

As can be seen, under visual presentation the mean RT for LAHV different matches was significantly longer than for HALV different matches ($p<.001$), while under auditory presentation the opposite was obtained ($p<.001$).

In a second experiment the modality of both the first and the second stimulus was varied (visual or auditory), as well as the similarity (HALV or LAHV) and the SOA (160 or 3000 ms; however, the article is not clear about whether this is SOA or ISI). The results indicated that the effects of similarity varied with the modality of the second stimulus regardless of first-stimulus modality or SOA-size. Furthermore, visual coding was used at both the long and the short SOA, indicating that in this study visual coding was as effective as auditory coding. This seems to conflict with Posner's conception of the development and use of the phonetic code (described above).

In his 1977-article Wood extended the design of the former experiment to three blocked conditions: A-A and V-V (both with an SOA of 1000 ms), and AV (with an SOA of 0 ms). The following results were obtained.

Stimulus condition:		A-A	V-V	AV	
Similarity:	HALV	509	452	622	(different response)
	LAHV	442	484	593	(different response)
	congruent	471	473	607	(same response)

Thus, matches seemed to be made with auditory codes in the A-A condition, but with visual codes in the V-V condition. In the AV condition recoding towards the auditory seems to occur ($p<.05$). One reason offered by Wood for matching on the auditory dimension was that V-to-A recoding was simply faster than A-to-V recoding. The first might be faster because the entire letter is available simultaneously with visual presentation, whereas the sequential nature of auditory presentation may require a delay of 200-300 ms. Other possible reasons were that a general propensity towards auditory coding prevailed in this particular experiment, or that subjects had more practice with visual-to-auditory recoding.

In a second experiment the first stimulus in a test pair was always presented auditorily, while the second stimulus was presented visually (A-V condition) or auditorily (A-A condition) at an SOA of 1000 ms. Inclusion of same pairs and non-test different pairs with the second stimulus distributed randomly between the visual and auditory modalities, led to 50% visual and 50% auditory second stimuli in both blocked conditions. The results indicated that matches were based on the modality dimension of the second stimulus even though the first stimulus was always auditory:

Stimulus condition:		A-A	A-V	
Similarity:	HALV	568	559	(different response)
	LAHV	552	576	(different response)
	congruent	587	596	(same response)

The set of results obtained by Wood (1974, 1977) shows that results of the matching task must be interpreted with care, since the code(s) used in the match may vary with the specificities of the task situation. It therefore seems useful to review converging evidence on cross-modal influences from other types of tasks.

2.2 Results of bimodal sublexical focused attention tasks

In a focused attention paradigm, Greenwald (1970a, 1970b, 1970c) presented simultaneous visual and auditory digits and recorded subjects' RTs to naming or writing the visual digits. In one experiment (1970a) the ISIs (between AV 200 to VA 200 ms) and Inter-Trial-Intervals (between 5 to 10 seconds) were varied between subjects because of technical reasons. In this experiment the RTs for the naming conditions were about 50 ms longer when the auditory digit was different from the visual than when both were name-identical (e.g., 400 vs. 350 ms). While the name-identical condition was generally faster than a no-auditory condition, the different condition was slightly slower over most ISIs (these conclusions are based on the observed RT-patterns in a figure in the article, since appropriate statistical tests are missing). A supplementary experiment used only an ISI of 0 seconds and spoken responses. The auditory stimuli were either (name-identical or different) digits or "taps" (the noise made by tapping the recorder microphone). Since RTs (measured from word onset) were fastest for name-identical digits and slowest for different digits, with the tap-condition in between, Greenwald concluded that there was evidence for response facilitation (in the name-identical condition) and interference (in the different condition). A second supplementary experiment showed that an auditory-irrelevant letter condition led to RTs about equal to the different-digit condition, but also to some more errors (7% vs. 5.7%).[4]

Mynatt (1977) tried to replicate and expand Greenwald's (1970b) results. In her first experiment subjects read visual digits accompanied by auditory name-identical or different digits, or bursts of 200 ms speech noise of similar intensity (mixed). Noise was used because studies had shown that RTs to visual stimuli are facilitated by any simultaneous auditory stimulus compared to presentation of the visual stimulus alone (e.g., Nickerson, 1970; Seif & Howard, 1975), and in order to approach the same signal-to-noise ratio in all three conditions. Subjects were asked to read the visual numbers as quickly as possible while ignoring the auditory stimuli. Analyses of variance on the median RTs showed significant main effects of Condition and of ISI (both $p<.001$). Though no post-hoc tests were reported, the accompanying figure shows that over a long range of ISIs (from 1 to 8 seconds) the name-identical condition was faster than the control condition, which itself was faster than the different condition. Longer ISIs led to longer RTs. It was concluded that the processing of an auditory "unattended" signal was not blocked, and that that signal's identity could differentially affect RTs to visual stimuli.

A second experiment used pairs of individual phonemes as stimuli. The task was to name letters by adding an /a/-sound to a presented letter consonant. Stop consonant letters (p, t, k, b, d, g) were paired with various auditory stimuli, leading to identical (e.g., p-p), shared-place (e.g., p-b), shared-voicing (b-d), and neither-shared (e.g., p-t) conditions. The shortest RTs were obtained for the identical condition, while the shared-voicing condition was faster than the shared-place condition. No difference was found between the combined two shared conditions and the neither-shared condition. RTs again increased considerably with the ISI going from 2 to 5 seconds. A third experiment replicated the second and added in a second block the consonants b, d, m, and n. Pairing these four consonants again led to identical (e.g., m-m), shared-place (e.g., b-m), shared-manner (e.g., m-n) and neither-shared (e.g., b-n) conditions. This experiment yielded the following pattern of results:

Shared features:					All	Place	Manner	None
Conditions:								
Old	(p,	t,	b,	d)	593	631	604	616
New	(b,	d,	m,	n)	580	596	593	586

The old conditions replicated the results of the second experiment. For the new conditions no significant results were obtained, but the identical condition showed a trend ($p<.10$). Post-hoc Scheffé-tests were significant for experiment 2 as a whole and for the old conditions in experiment 3.

In an attempt to explain the results, Mynatt proposed consonant-naming proceeded in two stages. In a first stage simultaneously presented stimuli are processed in parallel, with faster processing of voiced stimuli. The results are stored in short-term memory. In a second stage decision processes can start only when processing of both phonemes has finished. Identical phonemes lead to short RTs, because no decision about which stimulus is to be named is necessary (and less memory load occurs); different phonemes need a response decision, leading to longer RTs, the more so if the phonemes are voiceless.

However adequate this model may be, Mynatt's studies do not allow to decide whether the irrelevant auditory stimulus affects the visual information stream directly when the visual letter representation becomes available, or only later, when the naming response must be derived from that visual representation or when it is prepared for pronunciation (e.g., when the /a/ is added). This problem is rather severe here because the articulatory representation necessary for the naming response is often assumed to be similar to the phonological representation involved in auditory perception. Cross-modal effects at a response level are therefore not unlikely to occur.

While the experiments by Wood and Mynatt investigate the influence of different types of auditory accessories on responses to visual material, Gordon and Meyer (1984) present some results on cross-modal effects in the opposite direction. Though the focus of their article is on the relation between perception and production of phonetic features in speech, two experiments include a control condition that is of interest with respect to this thesis. In experiment 1 subjects had to memorize a set of syllable pairs constructed from the syllables puh, buh, tuh and duh (notation used by the authors). For example, one set included the pairs puh-puh, buh-duh, duh-tuh and tuh-buh. Paired syllables were matched with respect to voicing, and/or place of articulation, or neither. Afterwards, a choice-reaction task was run with an experimental condition, in which the subject had to produce the second (response) syllable of a pair vocally when the first (stimulus) syllable was auditorily presented; and a control condition, in which the stimulus syllable was presented visually instead of auditorally (e.g., as upper case "PUH"). RTs and error percentages for the control condition are presented in the Table below. The RTs of the voice+place (identical) condition differed significantly from the other RTs. Responses in the control condition exhibited no significant effects of either matched voicing or matched place of articulation when the syllables began with different consonants.

	Experiment 1		Experiment 5	
Shared features	RTs	Error	RTs	Error
voice + place	548	1.5	505	1.3
voice	660	5.4	645	8.6
place	666	6.2	645	7.5
none	661	6.4	659	6.1

In experiment 5 different syllables were used with fricative and sibilant consonants instead of stops: fuh, vuh, suh and zuh. For visual stimulus syllables, the shortest RTs again occurred when the consonants of the response syllables had both the same voicing and the same place of articulation as those of the stimulus syllables (i.e., when they were identical). As can be seen in the Table above, syllable pairs that had consonants with one matched feature and one mismatched feature yielded somewhat shorter RTs than pairs in which there were no matched features. However, these latter facilitation effects were unreliable. Also, subjects made 2% more errors in those conditions, suggesting a speed-accuracy trade-off. As in Mynatt's study, the identity effects in Gordon and Meyer's experiments cannot be allocated with certainty to early contacts of visual and auditory (perceptual) representations: They may have arisen during retrieval of the response syllable or during some articulatory stage.

Even though the direction of the cross-modal effects investigated differed between Mynatt's and Gordon and Meyer's experiments, a comparison of their results is interesting. In both types of experiments, the largest effects were found when the visual and auditory stimuli were nominally identical. This suggests that the cross-modal effects took place at the level of grapheme and phoneme representations, and not at that of smaller units. Of course, this requires an explanation of the small voicing effects in Mynatt's experiments, for example in terms of a late effect on the pronunciation (as suggested above).

2.3 Conclusion

In Chapter 1, five questions were posed with respect to the relationship of visual and auditory sublexical processing. The first question was whether grapheme and phoneme representations can cross-modally activate each other. The experimental literature just reviewed as a whole answers this question with "yes". Because of the problems with the (all-visual) matching paradigm, the best support for this answer is provided by the described focused attention experiments.

If cross-modal activation effects can occur, the next question is whether they occur both from the visual to the auditory modality, and in the other direction. In the literature there some evidence in favor of such bidirectionality, which comes, however, from only a few studies that applied a limited number of experimental techniques.

Most of the available research has not distinguished different possible temporal loci of cross-modal effects: early (e.g., during perception) or late (e.g., during articulation). Indeed, little is known about the time-course of cross-modal activation effects in general. The tasks used so far were not well-suited to investigate this (third) issue. The results of the successive matching task, for example, can reflect temporal aspects of cross-modal activation only indirectly, because when the second stimulus necessary for the match arrives, the first has been processed for some time and may thus already have exerted some cross-modal activation.

Furthermore, no systematic effort is made in the available research to distinguish facilitatory cross-modal activation effects between graphemes and phonemes from inhibitory effects. It is often not clear whether the RT-difference between two particular experimental conditions is due to facilitation, inhibition, or a combination of both. Therefore, the fourth question posed in Chapter 1 remains to be answered.

Not systematic answer is given to the fifth question either, whether cross-modal effects arise automatically, i.e. quickly and without conscious control of the subject. Since matching studies require a format common to the visual and auditory modality, they cannot decide whether any observed cross-modal effects are automatic or not. The naming studies by Mynatt (1977), and Gordon and Meyer (1984) provide some tentative evidence in line with automatic cross-modal effects.

Finally, the research reviewed has not explicitly considered to which extent the obtained cross-modal activation effects are due to general effects of visual and auditory stimuli or are specifically related to representational aspects of those stimuli.

To conclude, our knowledge about sublexical cross-modal activation effects is still rather incomplete. In the next three chapters of this thesis, seven experiments are reported that attempt to clarify the issues just considered.

3

Grapheme context effects on phonemic processing*

The processes underlying listening and reading necessarily differ in their early stages, due to differences in sensory organs and input signals. Nevertheless, since the same message can be understood in both modalities, the two processes must also converge at some point. In fact, most researchers would agree that reading and listening share the same database, the mental lexicon. This chapter examines a more controversial possibility, that of contact between visual and auditory representations at a lower level of representation. The smallest representational units for which such contact would seem feasible, are those of letters (graphemes) and speech sounds (phonemes).

Grapheme-phoneme contacts play an important role in several models of visual word recognition, such as the dual-route model (e.g., Coltheart, 1978, 1980). The dual-route model assumes that word recognition may proceed via either a direct or an indirect processing route. The first operates by mapping a word's extracted visual features directly onto its stored lexical representation; the second by translating the word's orthographic code into a phonological code and by subsequently using this phonological code to access the lexicon. It has been suggested (Coltheart, 1978; Venezky, 1970) that readers make use of an internal set of grapheme-phoneme correspondences (GPCs) that converts single letters or letter clusters into the corresponding phonemes or phoneme clusters. To account for the fact that effects of phonological recoding are found only under specific conditions (e.g., when pronouncing low-frequency words, Seidenberg, 1985a), the dual-route model assumes that the direct route usually wins the race to word recognition.

The assumption of a strong version of a dual-route theory that the indirect route alone suffices for word recognition, has been heavily criticized (e.g., Humphreys & Evett, 1985). Indeed this route runs into serious difficulties because there are many exceptions to the GPC-rules. Many correspondences between spoken and written words are inconsistent (e.g., the pronunciation of -AVE in HAVE) and even arbitrary (e.g., -OLO in COLONEL). This is due in part to the fact that writing systems reflect not only phonological principles, but also morphological and etymological ones.

The time-course model of visual word recognition (Seidenberg, 1985b, 1985c, 1987; Seidenberg, Waters, Barnes, & Tanenhaus, 1984) circumvents this problem by abandoning the assumption that the two routes function independently. In an implemented version of this model (Seidenberg & McClelland, 1989) orthographic representations are not "translated" or "recoded" into a phonological format, but activation is spread from orthographic to associated phonological representations. The additional time required to build up activation of phonological as compared to orthographic representations is taken to explain the relatively small influence that the phonological characteristics of a word have on visual word recognition. The specific patterns of connections between letter and sound strings are established via an associative learning process. This process leads to general rule-like behavior, while allowing irregularly spelled words to be learned and recognized.

The theoretical assumption that phonological information is activated during visual word recognition is empirically corroborated by experiments that rely upon the visual presentation of word and nonword stimuli. A number of lexical decision studies (Rubinstein, Lewis, & Rubinstein, 1971; Coltheart, Davelaar, Jonasson, & Besner, 1977), for example, have shown reaction time (RT) differences between nonwords that are homophonic with real words ("pseudohomophones"; e.g., BRANE) and nonhomophonic nonwords (BRAME). It is argued that pseudohomophones take longer to reject as nonwords because the homophonic words become activated via their nonlexically assembled phonological code.

Humphreys and Evett (1985) signal two problems with this conclusion. First, the effects may be due to orthographic rather than to phonological similarity of nonwords to words, since most studies have not satisfactorily partialed out these two variables. Second, it has not been established that phonological representations of nonwords are assembled without the use of lexical knowledge (e.g., Marcel, 1980).

Some recent studies have investigated the issue of phonological activation in visual word recognition with tasks other than lexical decision. Van Orden and his colleagues (Van Orden, 1987; Van Orden, Johnston, & Hale, 1988) had subjects decide whether a visually presented word belonged to a prespecified semantic category. Errors occurred more frequently with homophones of category instances (e.g., ROWS, homophonic to ROSE of the category "A FLOWER") than to spelling controls (ROBS). An elevated error score virtually identical to that of homophones was found for matched nonword homophone foils (such as ROAS). These findings demonstrate that the phonological properties of words mediate their recognition in a categorization task that requires reading for meaning. They further suggest that the phonological activation occurred automatically, since such activation was actually detrimental to performance in this situation.

A different approach was taken by Perfetti, Bell, and Delaney (1988), who asked subjects to identify briefly presented lower-case target words that were followed first by an upper-case pseudoword mask and subsequently by a pattern mask (a row of X's). They varied the orthographic and phonological properties shared by the target word and the pseudoword mask. When homophonic (MAYD) and orthographically similar (MARD) masks were equated for number of letters shared with the word target (made), both conditions led to a higher percentage of correct identifications of the target than a control mask, but an additional improvement in performance was found for the homophonic mask over the orthographically similar mask. The authors ascribed this effect to "phonetic activation"[1] and concluded that such activation occurred automatically (nonoptionally).

They further argued that the effects arose before word recognition, assuming that the process of target identification, still in progress at the onset of the mask, could be influenced by the mask's orthographic and phonological properties. In the authors' view the extent to which different types of information (graphemic, phonemic, categorical) contributed to the identification of a word would depend on the exact timing of activation patterns. Critical for the interpretation of visual word recognition experiments that manipulate the orthographic and phonological similarity of word and nonword stimuli is, therefore, the disentanglement of lexically and nonlexically mediated phonological effects, and of orthographic and phonological factors.

In contrast, bimodal studies have demonstrated that it is possible to test whether orthographic representations indeed activate associated phonological representations without these various confounding factors (e.g., Hanson, 1981; Kirsner, Milech, & Standen, 1983). In one such study, Frost, Repp, and Katz (1988) had subjects detect the presence of speech (either words or nonwords) in an auditory signal consisting of speech plus noise or noise alone. Simultaneous with the onset of the auditory stimulus, a matching visual stimulus (word or regular nonword), a nonmatching stimulus (word or nonword with similar structure), or a neutral stimulus (a row of X's) appeared. When amplitude-modulated masking noise was used, matching visual stimuli biased the subjects to report that speech was presented but did not improve the detectability of speech in noise. However, the RTs for correct word detections were facilitated with respect to the neutral and the no-match conditions. In an experiment involving nonwords much smaller bias effects were found, and RTs for correct detections were not faster with a visual-auditory match. On the basis of these results, the authors suggested that printed words were immediately and automatically recoded into an internal phonetic form. Since the effects for nonwords were much weaker, they proposed that the influence of the visual stimulus on speech processing was lexically mediated.

This last conclusion is at variance with the results of the visual experiments reported above that suggest cross-modal activation occurs before word recognition (cf. also Gordon & Meyer, 1984, Exps. 1 and 5). Possibly this specific task, involving only the detection of speech, was not sensitive enough to pick up such influences. To better understand cross-modal sublexical activation, we examined the influence of matching and non-matching visual context on speech processing with a bimodal task that required not just detection but identification of the auditory stimulus. By varying not only the structural but also the temporal relationship between the visual and auditory stimuli we hoped to find out *how* and *when* representations of letters influence those of speech sounds.

More specifically, Dutch subjects made a forced two-choice decision on the identity of target vowels (e.g., /a:/ or /e:/) in auditory syllables (e.g., /ta:/ or /ke:/). Before, at, or after the onset of the auditory syllable, visual letter primes of two types could appear: indirect or direct primes. In *indirect* priming the letter was nominally congruent or incongruent with the consonant appearing in the same syllable as the target vowel (e.g., letter P, syllable /pa:/ with target vowel /a:/). Indirect priming effects on the vowel would indicate that cross-modal influences occur even when the visual stimulus is neither directly task-relevant nor connected to a response. In other words, such effects could be considered "automatic" (Posner &

Snyder, 1975a, b). In contrast to indirect priming in which the letter was congruent or incongruent with the consonant in the target syllable, in *direct* priming the letter (e.g., A) was congruent or incongruent with the target vowel itself (e.g., letter A with syllable /a:/, where /a:/ indicates pronunciation of the letter A in Dutch). Once the existence of a connection between representations for a letter and a corresponding target vowel is shown, this direct priming technique should allow us to study the temporal development of the cross-modal influence. In sum, we first investigated the existence of automatic cross-modal effects in two experiments using the indirect form priming technique. We subsequently explored temporal aspects of cross-modal influences in more depth in a third experiment involving direct as well as indirect priming.

3.1 Experiment 1

In Experiment 1 the visual stimulus was always a letter that was either name-congruent or -incongruent with the consonant of an auditory CV-syllable (e.g., letter P or K, syllable /pa/). We expected that the congruence or incongruence between the letter and consonant would influence the response to the target vowel. This would be the case if the following assumptions hold. First, a graphemic representation of a letter is connected to its corresponding phonemic representation. Second, activation of the graphemic representation spreads to the corresponding phoneme. Third, more efficient processing of the consonant in the auditory CV-syllable leads to faster identification of the vowel in that syllable. The first two assumptions are compatible with several connectionist models of word recognition (e.g., variants of the time-course model of visual word recognition by Seidenberg et al., 1984). We expected the third assumption to hold on the basis of results like those of Wood and Day (1975) showing that the phonetic information corresponding to the consonantal and vocalic parts of such CV-syllables is processed as an integral unit. These authors showed that speeded phonetic decisions about vowels were influenced by the preceding stop (and vice-versa). In our experimental situation, the plosive consonant should be activated by the congruent letter, but not by the incongruent one; the more activated consonant should lead to faster identification of the following vowel. Hence, a facilitatory influence of the visual letter information upon the processing of the congruent consonant should also be observed upon the processing of the vowel.[2]

In sum, these three assumptions led us to predict faster RTs in a consonant-congruent condition (like P-/pa/) than in a consonant-incongruent condition (like K-/pa/). Furthermore, by varying the stimulus onset asynchrony (SOA) of the prime and the target syllable, we hoped to determine the time range over which the expected cross-modal influence could be obtained. Finally, we reasoned that any results showing priming should reflect automatic cross-modal activation, since in the indirect priming procedure used, the letter's identity is irrelevant to the forced-choice response to the vowel and consequently is unlikely to induce conscious strategies in the subject.

3.1.1 Method

Subjects. Thirty undergraduates at Nijmegen University, all native speakers of Dutch, were paid to participate in the experiment.

Design and Stimuli. Bimodal test stimuli were used consisting of a visually presented letter and an auditorily presented nonword syllable. The letter was P, T, or K. The six different syllables consisted of a consonant (/p/, /t/, or /k/) and a short vowel (/a/ or /o/). All letters were combined with all syllables, leading to 18 different bimodal test stimuli. Depending on the relation between the letter and the syllable, a bimodal stimulus belonged to one of two experimental conditions: a consonant-congruent condition, in which the letter was nominally identical to the consonant in the auditory syllable (e.g., letter P paired with syllable /pa/); or a consonant-incongruent condition in which the letter and the consonant differed (e.g., letter K and syllable /pa/).

The temporal relationship between the visual and auditory stimuli was varied: the stimulus onset asynchrony (SOA) was −190, −70, −30, +30, or +150 ms. For example, at SOA1 the visual stimulus was presented 190 ms before the onset of the auditory syllable, while at SOA5 the visual stimulus was presented 150 ms after the onset of the auditory syllable.

The six stimuli in the congruent condition were repeated 10 times under each SOA, while the 12 stimuli in the incongruent condition were repeated 5 times under each of the five SOAs. To keep the subject's attention directed to the screen, 80 catch trials (in which the subjects should not respond) were constructed in which the visual letter H was combined with all target syllables under the first three SOAs.

The resulting 6*10*5 (congruent) + 12*5*5 (incongruent) + 80 (catch) = 680 experimental trials were randomized and divided into two blocks of 340 trials each. Furthermore, 20 practice trials were added, bringing the total number of trials in the experiment to 700.

The auditory stimuli were recorded on tape by a female native speaker of Dutch in a sound-proof room. The length of the stimuli varied from 180 to 200 ms. The stimuli were digitized on a VAX 11/750 computer with a sampling rate of 20 kHz. and a randomized sequence of targets was placed on one channel of a tape. The output of the computer was low-pass filtered with a cutoff frequency of 10 kHz. During the experiment the auditory stimuli were presented binaurally over headphones. On the second channel of the tape a pulse, inaudible to the subjects, was placed, that triggered both the timer for the recording of the response latencies and the presentation of the visual stimulus (after a delay that depended on the specific SOA in a trial).

The visual stimuli were white Roman capitals, 6 mm in height, presented on a MATROX-screen with a dark background. The monitor was placed at a distance of 60 cm from the subject, in order to provide projection within the foveal field of the eye. The visual stimuli were presented for 60 ms. Presentation of the visual stimuli and recording of the reaction times were controlled by a PDP-11/23 computer.

Procedure. Subjects read the written instructions, which were repeated orally at the beginning of the experiment. They were instructed to rest their two index fingers lightly on the two response buttons in front of them, and to push the /a/-button as fast as possible whenever they heard the /a/-vowel, and push the /o/-button when the heard the /o/-vowel. The /o/-response button was allocated to the index finger of the preferred hand. Subjects were also told that sometime before or during the presentation of the auditory syllable, a letter would appear on the screen, and that they were not to respond if the letter was H.

Each trial consisted of a 1000 Hz. warning signal of 150 ms duration, followed after 500 ms of silence by the auditory syllable. A variable interval after the warning signal (depending on the SOA in the trial) a visual letter stimulus appeared. A new trial was initiated every 5.5 seconds.

The session consisted of 20 practice trials followed by two blocks of 340 test trials. After the practice set there was a short pause in which subjects could ask for clarifications. Between the two experimental blocks there was a pause of three minutes. In all, the experiment took about one hour and fifteen minutes.

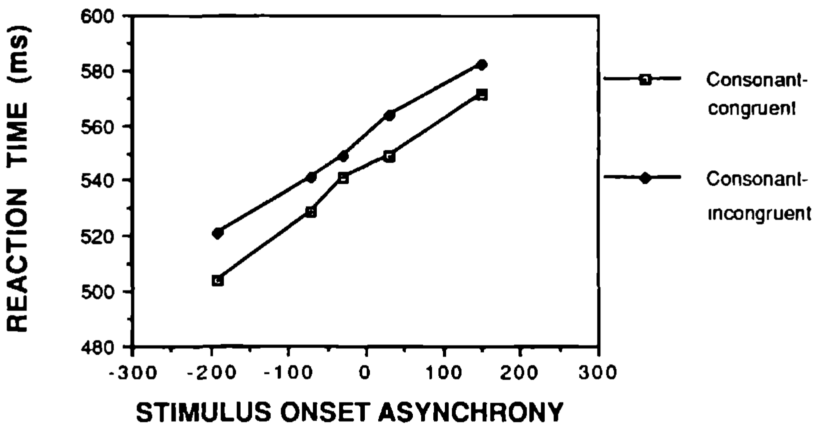


Figure 3.1. Mean reaction times (in ms) for consonant-congruent and -incongruent conditions as a function of SOA.

3.1.2 Results

Mean RTs (measured from the onset of the auditory syllable) were computed for each subject and for congruent and incongruent conditions at each SOA. The percentage of missing RTs and RTs greater than 1000 ms or smaller than 150 ms was 6.5, and was distributed equally across the congruent and incongruent conditions. Missing values were substituted by mean RTs in each relevant subcondition for each subject. The percentage of errors on the H-trials (false alarms) was 3.6. Figure 3.1 graphically displays the RT-pattern of results for the congruent and incongruent conditions as a function of SOA. The exact values of the mean RTs can be found in Appendix 2.

A by-subject analysis of variance with the factors Congruence (congruent vs. incongruent), Syllable Type (with initial consonants /p/, /t/ or /k/) and SOA showed significant main effects for all these factors (Congruence [$F(1,29)=26.80$, $p<.001$]; Syllable Type [$F(2,58)=11.42$, $p<.001$]; and SOA [$F(4,116)= 40.96$, $p<.001$]. A significant interaction was found between Congruence and Syllable Type [$F(2,58)=5.71$, $p<.01$]. No significant interactions of any factors with SOA were obtained.

Paired planned comparisons of the congruent and the incongruent conditions for each of the three syllable types showed significant differences for syllables with /p/- and /t/-consonants, but not for those with /k/-consonants, as indicated in Table 3.1. A further analysis of the syllables containing a /k/-consonant showed that the RT-difference between congruent and incongruent conditions was significant for the syllable /ka/ (535 vs. 548 ms, $t(29)=2.68$, $p=.01$) but not for /ko/ (563 vs. 557 ms, $t(29)=.98$, $p>.10$).

condition	congruent	incongruent	t(29)	p-value
/p/-syllable	539	554	4.61	p<.001
/t/-syllable	528	547	4.64	p<.001
/k/-syllable	549	552	.95	ns

Table 3.1. Mean reaction times (in ms) and planned comparisons of means for /p/-, /t/- and /k/-syllable types in consonant-congruent and -incongruent conditions.

3.1.3 Discussion

The most important finding of this experiment was a significant RT-advantage of the consonant-congruent condition over the -incongruent condition. This congruence effect can best be understood as being the result of cross-modal activation of an auditory consonant representation by a visual letter representation. To obtain these results in an indirect priming procedure where grapheme-phoneme congruence by itself has no predictive value for the vowel response suggests that the cross-modal activation is *automatic.* Indeed, for choosing the correct vowel a cross-modal comparison is neither required nor directly relevant. Consequently, these results provide evidence for automatic activation of phoneme representations by graphemes.

The absence of an interaction between the factors, Congruence and SOA, raises some questions about the temporal characteristics of this cross-modal activation of the consonant by the letter prime. One might have expected the influence of the letter prime upon the processing of the vowel to decrease or even to disappear at later SOAs, but it appears that this facilitatory influence remained roughly constant across the entire SOA range. However, there was still an interval of approximately 400 ms between the onset of the letter (in the latest SOA-condition) and the response, during which the cross-modal effect could have taken place.

The observed gradual increase in RT over SOA for consonant-congruent and -incongruent conditions could be a consequence of the catch-trial condition with H. In order to avoid reacting in H-trials, subjects might delay their response until the letter is at least partially identified. Thus, the later the letter appears in time relative to the auditory syllable, the longer the subject would have to wait. This potentially negative consequence of the H-condition led us to replicate the experiment with some changes in design.

3.2 Experiment 2

In Experiment 2, the catch-trial condition was eliminated and the subject only made a decision on the vowel in the syllable. To ensure that the subject paid attention to the visual stimuli, an additional off-line task was used: After a varying number of trials, the word "RAPPORT" ("report") appeared, indicating that the subject had to write down whether a letter was presented on the last trial or not.[3] To make this off-line decision non-trivial, Experiment 2 included auditory single-channel trials. This single-channel condition served also as a baseline for the evaluation of congruence effects. Finally, we also attempted to determine whether the differences between consonant-congruent and -incongruent conditions generalized to other syllable types. For this purpose, fricative - vowel syllables were included as it has been shown that the cues to their component phonemes are processed interdependently (Tomiak, Mullennix, & Sawusch, 1987; Whalen, 1984).

3.2.1 Method

Subjects. Twenty-seven undergraduates at Nijmegen University, all native speakers of Dutch, were paid to participate in this experiment.

Design and Stimuli. The design of this experiment was similar to that of Experiment 1. However, in addition to the two experimental conditions of Experiment 1, a single-channel condition was included. Bimodal stimuli consisted of a visually presented letter (P, K or S) and an auditorily presented CV-syllable. The consonant of the syllable was /p/, /k/ or /s/; the vowel was /a:/ or /e:/. The choice of long vowels in this experiment was motivated by their predominance in free-occurring Dutch CV-syllables.[4] The length of the six syllables varied from 430 to 500 ms. The single-channel conditions consisted of syllable presentation without a letter. All visual stimuli were presented for 100 ms.

Stimuli in the congruent condition (e.g., letter S combined with syllable /se:/) were repeated 10 times under each of five SOAs, while stimuli in the incongruent condition (e.g., letter K with syllable /sa:/) were repeated five times under each SOA. The SOAs were −190, −70, −30, +30, and +150 ms. In addition, there were 30 single-channel trials of each of the six auditory syllables. Thus, the experiment consisted of 6*10*5 (congruent) + 12*5*5 (incongruent) + 180 (single-channel) = 780 test stimuli. These were randomized and divided into two sessions of 390 trials each. Furthermore, for each session 60 report trials, which consisted of the visual presentation of the word "RAPPORT", and 47 practice trials were constructed. In all, each experimental session included 497 trials.

Procedure. The same two-choice procedure was used as in the previous experiment. Subjects participated in the two sessions on successive days. The order of sessions was counterbalanced over subjects. Response allocation across the two sessions was also counterbalanced over subjects. Half of the subjects reacted to /e:/ with their right index finger on the right response button in the first session, and with the left index finger on the left response button in their second session the next day. The other half of the subjects was instructed to do just the opposite. In addition, whenever the word "RAPPORT" was visually presented, subjects were to indicate on a prestructured form whether the preceding trial had included a visual stimulus or not. A 4-s interval followed each trial, except report trials, which were followed by a pause of 6 seconds. Each experimental session lasted for 40 minutes with a short break after about 20 minutes.

3.2.2 Results

Mean RTs (measured from the onset of the auditory syllable) were computed for each subject and each experimental condition. The percentage of missing RTs and RTs greater than 1000 ms or smaller than 150 ms was 4.3, and was distributed equally across consonant-congruent and -incongruent conditions. Missing values were substituted by mean RTs in each subcondition for each subject. The percentage of errors in the subjects' judgement

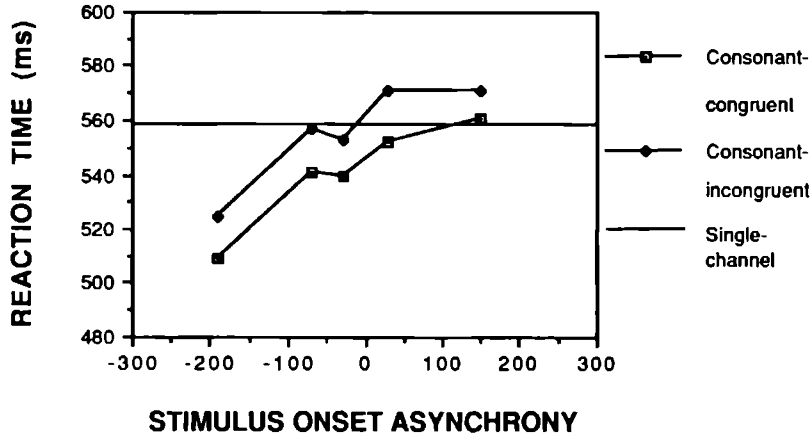


Figure 3.2. Mean reaction times (in ms) for the single-channel condition, and for the consonant-congruent and consonant-incongruent conditions as a function of SOA.

about whether a letter appeared in the trial preceeding the RAPPORT-trials was 5.6. The RT-patterns over SOA for the congruent and incongruent conditions as well as the mean RT for the single-channel condition are shown in Figure 3.2 (for the exact RT-values see Appendix 2).

An analysis of variance with the factors Congruence (consonant-congruent vs. -incongruent), SOA, and Syllable Type (with initial consonants /p/, /k/ or /s/) showed main effects of Congruence [$F(1,26)=39.82$, $p<.001$], SOA [$F(4,104)= 42.06$, $p<.0001$], and Syllable Type [$F(2,52)=95.50$, $p<.0001$]. A significant interaction was found between Congruence, SOA and Syllable Type [$F(8,208) = 3.83$, $p<.001$].[5] The other interactions were not significant.

3.2.3 Discussion

Both Experiment 1 and 2 showed faster RTs in the consonant-congruent than in the consonant-incongruent conditions, despite several differences in the procedure used. Congruence effects were found for syllables with stop consonants (Experiments 1 and 2) and with fricative consonants (Experiment 2); and when the task explicitly demanded on-line identification of the visual stimulus (Experiment 1), and when it did not (Experiment 2). Furthermore, both experiments showed a congruence effect over the entire SOA range from −190 (letter first) to +150 (syllable first), showing that even when the auditory syllable leads by 150 ms, response time to the vowel is influenced by the congruence or incongruence of letter and consonant.

The existence of this congruence effect provides strong evidence for sublexical activation of auditory consonant representations by congruent visual letters. Since the relationship between the auditory consonant and the letter is not relevant for the subject's vowel decision, we can conclude that the activation takes place automatically, that is, without conscious control by the subject (Posner & Snyder, 1975a, b). This demonstration of cross-modal activation with the indirect priming procedure confirms and extends the results of other studies such as those using bimodal same/different matching (cf. Wood, 1977; Posner, 1978), where a direct comparison between graphemes and phonemes is required by the task itself.

Experiment 1 and 2 both revealed a gradual increase in the RTs for all conditions with increasing SOA. One explanation proposed for this effect in Experiment 1 depended upon the presence of the catch-trials. It was argued that subjects delayed their response until they had partially identified the visual stimulus, because they were not to respond when the letter H appeared. However, since Experiment 2 showed an increase of RT over SOA even without this catch-trial condition, we can eliminate this explanation.

According to another explanation for the increase of RT over SOA, the arrival of the visual stimulus causes a shift of attention away from the auditory modality and disrupts processing (cf. Miller, 1985, p. 520). If this attention shift occurs while the auditory stimulus is being processed (late letter presentation), it should lead to greater disruptions and stronger increases in RT than when the auditory stimulus has not yet arrived (early letter presentation). This view, in which the visual stimulus has a detrimental effect, may be contrasted with one in which the visual stimulus functions as an alerting cue, leading to preparation-enhancement (cf. Nickerson, 1973). The amount of facilitation caused by the visual stimulus would depend on its relative time of arrival with respect to the auditory target stimulus: Early visual stimuli would facilitate the RT more than late. If the single-channel condition is taken as a baseline, the increase in

RT over SOAs is better explained by the preparation-enhancement hypothesis, since for most SOAs the bimodal conditions show a facilitatory effect (see Figure 3.2).

3.3 Experiment 3

The previous experiments using the indirect priming technique have not allowed us to determine the time-course of phoneme activation by graphemes, since there was no interaction between the size of the congruence effect and the SOA. Furthermore, the interpretation of the results in terms of the timing of the cross-modal influence is complicated here because it is difficult to disentangle the temporal properties of this influence from those of the processing of the vowel. As a consequence, we decided to prime the auditory vowel directly in order to obtain a more direct reflection of the spread of activation from the visual prime to the auditory target under different SOAs. In the next experiment we therefore added conditions in which the letter prime was congruent with the target vowel or not (e.g., in Dutch, A-/ka:/ vs. A-/ke:/).

As a baseline condition for each SOA, a bimodal condition was included in which a nonlinguistic visual stimulus accompanied the auditory syllable. Effects of attention shift or alerting caused by the presence of a visual stimulus should be similar for baseline and test conditions. As in Experiment 2, an auditory single-channel condition was also included to examine the general effect of adding a visual accessory to the auditory syllable.

In addition to CV-syllables (like /ka:/), Experiment 3 also included syllables with two other structures: VC-syllables (like /a:k/) and V-syllables (like /a:/). If the integration account described above is correct, subjects would integrate the consonantal information with that of the target vowel in the case of VC-syllables. Thus, the RTs to VC-syllables should be longer than those to V-syllables, provided that the former are longer than the latter and that the duration of the vowel in the VC-syllable is shorter than the V-syllable.

3.3.1 Method

Subjects. Twenty-eight undergraduates at Nijmegen University, all native speakers of Dutch, were paid to participate in the experiment.

Design and Stimuli. Visually presented stimuli were the letters K, P, A and E and the symbol *; auditory stimuli were the syllables /ka:/, /ke:/, /a:/, /e:/, /a:k/ and /e:k/.[6] The naturally pronounced syllables were matched in length as much as possible: the duration of /ka:/, /ke:/, /a:/ and /e:/ was about 365 ms; the duration of /a:k/ was about 465 ms, that of /e:k/ 495 ms. The length of the vowel in /a:k/ and /e:k/ was about 215 ms. All visual stimuli were presented for 100 ms.

Each visual stimulus was combined with each syllable. In the consonant-congruent conditions the letter was nominally identical to the consonant (letter K combined with syllables /ka:/, /ke:/, /a:k/ or /e:k/), in the consonant-incongruent conditions it was not (letter P combined with these syllables). In the vowel-congruent conditions the letter was nominally congruent to the target vowel (e.g., letter A combined with /ka:/, /a:/ or /a:k/), in the vowel-incongruent conditions it was not (e.g., letter E combined with syllables just mentioned). Furthermore, there was a condition in which a letter consonant was combined with a V-syllable (e.g., P-/e:/). In the star-condition the auditory syllable was presented with a visual *. A single-channel condition was also included, in which the syllable was presented in isolation.

As before, the temporal relationship between the visual and auditory stimulus in bimodal conditions was varied. However, in order to examine possible cross-modal effects over a longer time range, the SOAs used in this experiment were −250, −100, 0, +100 and +250 ms. A negative number indicates that the visual context preceded the auditory syllable.

Each bimodal stimulus was repeated 10 times under each SOA. Each single-channel stimulus was repeated 20 times. Therefore, the number of test stimuli in the experiment was 6 (syllables)* 5 (visual stimuli)* 5 (SOAs) * 10 (repetitions) + 6 (syllables)*20 (repetitions) = 1620. These were randomized and divided into three experimental sessions such that each session consisted of 540 trials. Furthermore, 60 report trials (consisting of the visual presentation of the word "RAPPORT") and 40 practice trials were constructed for each session. Thus, subjects were presented with 640 trials per session.

Procedure. Subjects participated in three sessions on successive days. The order of sessions was counterbalanced over subjects. Half of the subjects reacted to /e:/ with their right index finger, and to /a:/ with the left index finger. The other half of the subjects was instructed to do just the opposite. Thus handedness and response button were counterbalanced. In

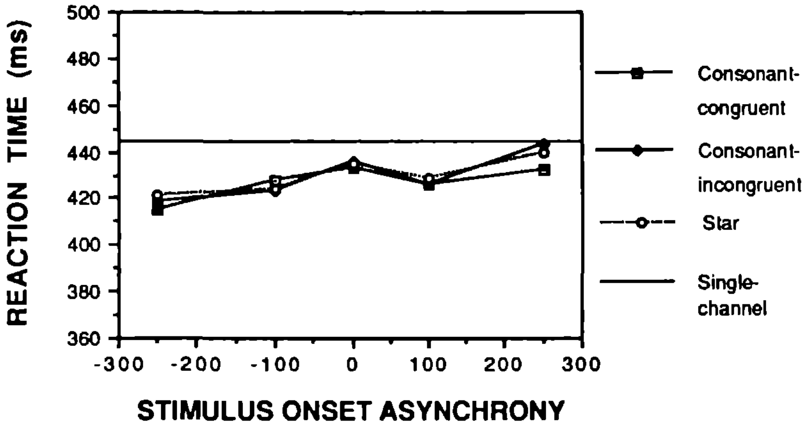


Figure 3.3. Mean reaction times (in ms) for the single-channel condition, and for the consonant-congruent, -incongruent, and star-conditions as a function of SOA.

other respects the procedure was the same as in the previous experiment. Each experimental session lasted for about an hour and had a short break after about 35 minutes.

3.3.2 Results

Mean RTs (measured from the onset of the auditory syllable) were computed for each subject and each experimental condition. The percentage of missing RTs and RTs greater than 1000 ms or smaller than 150 ms was 1.2 for the double channel conditions and 2 for the single-channel condition. Missing values were substituted by mean RTs in each condition. The percentage of errors in the subjects' response on the RAPPORT-trials was 6.0.

The data were analysed separately for the two types of congruence. First results for all consonant-congruent and -incongruent conditions involving CV- and VC-syllables are presented, then results for the vowel-congruent and -incongruent conditions. The corresponding star-conditions

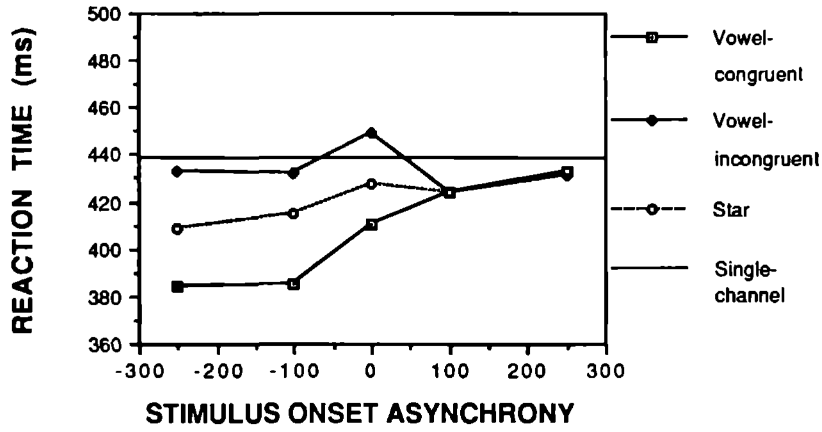


Figure 3.4. Mean reaction times (in ms) for the single-channel condition, and for the vowel-congruent, -incongruent and star-conditions as a function of SOA.

were included in the analyses.

Mean RTs for the Consonant-Congruence data are graphically displayed in Figure 3.3 (also see associated Table in Appendix 2). An analysis of variance on these data with the factors Congruence Type (consonant-congruent, consonant-incongruent, star) and SOA led to a main effect for SOA [$F(4,108) = 7.67$, $p<.0001$]. No significant main effect was found for Congruence Type [$F<1$], nor any significant interaction of Congruence Type and SOA [$F<1$].

Mean RTs for the Vowel-Congruence data are graphically displayed in Figure 3.4 (associated Table in Appendix 2). An analysis of variance on these data with the factors Congruence Type (vowel-congruent, vowel-incongruent, star) and SOA showed significant main effects for Congruence Type [$F(2,54)=49.34$, $p<.0001$], and SOA [$F(4,108)=19.84$, $p<.0001$], and a significant interaction effect for Congruence Type and SOA [$F(8,216)= 16.91$, $p<.0001$].

We further related the vowel-congruence and -incongruence conditions to the star- and single-channel conditions. Planned comparisons between the star-conditions and the vowel-congruent or -incongruent conditions for each SOA showed significant differences for all comparisons under the first three SOAs (cf. Figure 3.4). Planned comparisons between all these bimodal conditions and the single-channel condition showed significant differences for all comparisons under these SOAs as well, except for the vowel-incongruent condition, that did not differ significantly from the single-channel condition at SOA1 and SOA2 (cf. Figure 3.4). At SOA4 all bimodal conditions were significantly faster than the single-channel condition, but did not differ among each other. At SOA5 no significant differences were found any more.

In order to test if the three Syllable Categories (CV, V, and VC) behaved differently over SOA, an analysis of variance was run on the combined vowel-congruence data (vowel-congruent, vowel-incongruent, and star-conditions) with the factors Syllable Category and SOA. Mean RTs, including those of the single-channel condition, are graphically represented in Figure 3.5 (associated Table in Appendix 2). The analysis showed significant main effects for Syllable Category [$F(2,54)=95.75$, $p<.0001$], SOA [$F(4,104)=19.84$, $p<.0001$], and a significant interaction of Syllable Category and SOA [$F(8,216)=2.85$, $p<.01$].

Planned comparisons between the different Syllable Categories for each SOA showed that reactions were significantly faster to V-syllables than to CV-syllables at all five SOAs, but faster than to VC-syllables at only two SOAs (SOA2=-100 and SOA5=+250).

3.3.3 Discussion

Experiments 1 and 2 have demonstrated automatic cross-modal grapheme-to-phoneme activation. Experiment 3 provides information on how such activation develops across time. Large differences between vowel-congruent and -incongruent conditions were found when the letter appeared simultaneously with or before the auditory syllable. In order to distinguish facilitatory and inhibitory components of this vowel-congruence effect, these conditions were compared to the star-condition. Relative to this baseline, strong facilitation effects were obtained for the vowel-congruent condition and strong inhibition effects for the -incongruent condition, when the visual stimulus preceded the auditory syllable by 250 or 100 ms. Both effects disappeared completely when the letter was presented 100 or 250 ms after the onset of the syllable. The facilitation effect can be interpreted as evidence for cross-modal activation of the target vowel-representation by the

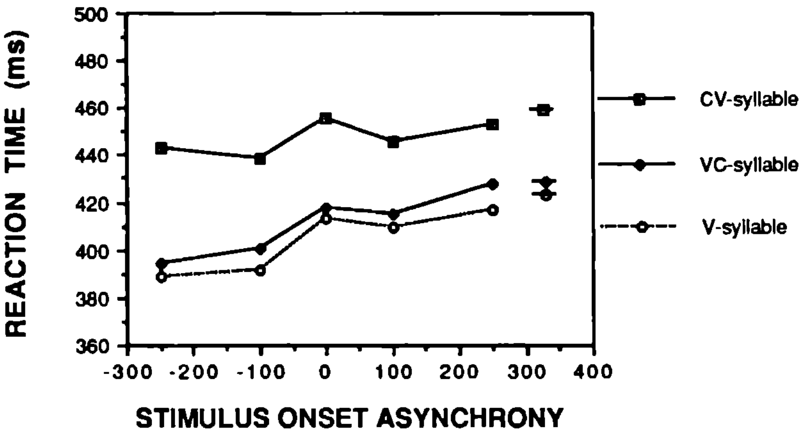


Figure 3.5. Mean reaction times (in ms) for CV-, V- and VC-syllable types in the single-channel condition, and in the combined vowel-congruent, -incongruent and star-conditions as a function of SOA.

letter stimulus, while the inhibition effect suggests cross-modal activation of the phonological representation of the vowel connected to the competing response.

Comparison of the data with the single-channel condition revealed a more global facilitatory influence of the visual stimulus. Even when the auditory syllable led by 100 ms (SOA4), and the RT-differences between vowel-congruent, -incongruent and star-conditions had disappeared, a significant facilitation remained in comparison with the single-channel condition. This effect disappeared only at an SOA of 250 ms. We attribute this effect to a general non-specific alerting of the subject by the visual stimulus, as we did for the decrease in RTs for the negative SOAs in the first two experiments.

The analysis of the RTs for the different syllable categories (CV, V, or VC) over the combined vowel-congruent, -incongruent, and star-conditions showed that RTs to CV-syllables were longer than those to V- and VC-syllables by 25 to 50 ms. Since RTs were measured from stimulus onset, this difference is best accounted for by the delayed arrival of the vowel in the CV-syllables compared to the V- or VC-syllables. Slightly longer RTs were found to VC-syllables than to V-syllables, indicating that the subjects took into consideration some information about the following consonant as was predicted by the account appealing to integration of consonant and vowel information (cf. Healy & Cutting, 1976).

The RT-differences between consonant-congruent and -incongruent conditions, robust in Experiments 1 and 2, disappeared in Experiment 3. To the extent that the literature on semantic priming is relevant here, we can find some explanations for this unexpected result. For example, it is known that the types of experimental conditions, as well as their relative proportion, influence the size of the semantic priming effects observed (e.g., Neely, 1977). The inclusion of the vowel-congruent and vowel-incongruent conditions in our Experiment 3 changed the proportion of consonant-congruent conditions, thus perhaps washing out the more subtle effects for the consonant-congruent and -incongruent conditions. Furthermore, the priming literature suggests that faster RTs generally lead to smaller priming effects (e.g., Flores d'Arcais, Schreuder, & Glazenborg, 1985). The fact that the RTs in Experiment 3 were about 100 ms faster than those in the first two experiments may have contributed to the disappearance of the consonant-congruence effect (cf. the range effects described by Poulton, 1973). Further experimentation is clearly needed to investigate how appropriate these explanations really are for our form priming experiments.

3.4 General discussion

In a series of three experiments we examined the cross-modal influence of graphemes on phonemes and the time-course of this influence. The existence of cross-modal activation was demonstrated in Experiment 1 and 2 using an indirect priming technique in which the priming visual letter stimuli were either congruent or incongruent with the consonant of an auditory syllable (e.g., letter P or K, syllable /pa:/). Over the entire range of SOAs used, a RT-advantage for the vowel target was found in the consonant-congruent over the -incongruent condition. To explain this advantage, we must assume automatic cross-modal activation of the representation of the consonant phoneme by the letter prime.

The temporal development of cross-modal sublexical activation was explored in more depth in Experiment 3 with a direct priming technique.

Here the letter stimuli presented were either congruent (e.g., letter A, syllable /ka:/) or incongruent (letter E, syllable /ka:/) with the target vowel in the auditory syllable. For trials in which the letter preceded the auditory syllable, strong facilitation and inhibition effects were obtained when the vowel-congruent and -incongruent conditions were compared to the bimodal baseline-condition. These effects decreased when the letter was presented at the onset of the auditory syllable, and disappeared when the letter was presented during the auditory stimulus. The evolution of the facilitatory and inhibitory effects over SOA suggests that it took time for the letter to activate its corresponding phonemic representation. When the letter came late, its contribution to the activation of this phonemic representation could no longer influence the response.

More specifically, the results of Experiment 3 support the following account of bimodal processing in this forced-choice task. In the bimodal conditions, the letter accompanying the auditory signal generates an orthographic representation which automatically activates an associated phonological representation. In the vowel-congruent condition with negative SOAs, the letter preactivates the same representation as is activated by the vowel, and this additional activation speeds up the response compared to the baseline condition. Indeed, in this baseline condition, the star symbol does not activate any phonological representation. When the phonological representation activated by the letter differs from that of the target vowel, its influence depends on whether it belongs to the response set or not. When it does not belong to the response set (e.g., the phonological representation corresponding to the letter P), it does not influence the response, producing no RT-difference between this condition and the baseline star-condition. However, when the phonological representation activated by the letter (e.g., E) does belong to the response set, the target vowel and the letter activate phonological representations associated with competing responses, leading to the inhibition found for the vowel-incongruent condition.

The explanation we have advanced for our results is similar to that proposed for the results obtained with a family of tasks often applied in the domain of visual word recognition, such as the Stroop task (e.g., Stroop, 1935; Glaser & Glaser, 1982), the picture-word naming task (e.g., Glaser & Düngelhoff, 1984; LaHeij, 1988) and the flanker task (e.g., Eriksen & Eriksen, 1974; Hell, 1987). In a typical Stroop task, color words (like RED) are printed in a color that is congruent or incongruent with their name (e.g., in red or blue). If subjects have to name the color in which the words are printed, naming time increases up to 100 ms when the print color and word meaning are incongruent, as compared to a situation where the color of a meaningless letter string is named. When this experimental situation is translated into that of Experiment 3, the target vowel corresponds to the color, and the visual letter to the orthographic form of the printed word.

As in the Stroop task, (vowel-)incongruent conditions led to interference effects, and (vowel-)congruent conditions to facilitation relative to a bimodal baseline condition. The development of the facilitation and inhibition effects over SOA is also comparable (cf. Glaser & Glaser, 1982; Glaser & Düngelhoff, 1984). With Van der Heijden (1981) we would like to suggest that these similarities across tasks reflect general principles in the mechanisms used to perform different types of attentional tasks. Since Stroop effects are usually interpreted to be indicative of automatic processing (cf. Kahneman & Chajczyk, 1983), the similarity of the current data and those obtained in Stroop tasks then implies the same for our congruence effects.

After having established the existence of automatic cross-modal activation of phonemes by graphemes, we must now assess the relevance of this conclusion for visual word recognition. If graphemic information automatically activates phonological representations in situations in which the visual stimulus serves mainly as an accessory and is not directly task-relevant, such cross-modal activation is even more likely to occur in visual word recognition, where attention must be paid to visual stimuli and where phonological information can play a useful role. We therefore would like to argue that our results, although obtained with nonword syllables, support the hypothesis that sublexical grapheme-to-phoneme activation occurs automatically in visual word recognition.

By measuring the activation of phonological representations by graphemes via an auditory task, we have not concerned ourselves directly with visual word recognition. Nonetheless, we have made certain that the representations activated are the same as those activated in auditory perception. This is not necessarily the case for visual tasks, since they do not directly measure phonological effects. There is no guarantee that the phonological effects arising for visually presented words or nonwords are truly "phonological". Thus, both approaches provide complementary evidence for grapheme-to-phoneme activation in word recognition. Taken together, the relevant experimental evidence obtained exclusively in the visual domain (Van Orden, 1987; Van Orden, Johnston, & Hale, 1988; Perfetti, Bell, & Delaney, 1988) and the bimodal results for words (e.g., Hanson, 1981) and nonwords (this work) offer convergent evidence for the existence of automatic, not lexically mediated, grapheme-to-phoneme activation during visual word recognition.

The activation metaphor we adopted in this chapter seems particularly well-suited to account for the decrease in facilitation and inhibition effects over SOA in Experiment 3. Indeed, all results are easily understood within the context of interactive activation models, such as the time-course model of visual word recognition (e.g., Seidenberg, 1985b). If the model were expanded to include the possibility of both orthographic and phonological input, the effects of visual context on auditory representations could be simulated. Furthermore, while our data are accounted for by this type of model quite well, they are hard to reconcile with all those models of visual word word recognition in which phonological information becomes available only after word recognition.

In sum, bimodal experiments of the type presented here may help to constrain models of visual word recognition by providing information about structural and temporal aspects of cross-modal activation. Our results indicate that such models should incorporate a mechanism of sublexical activation of phonemic representations by graphemes.

4

Bidirectional grapheme-phoneme activation in a bimodal detection task

"Assailed by eyes, ears, nose, skin, and entrails at once", we should experience the world "as one great blooming, buzzing confusion", as William James (1890) put it, were we not able to integrate information arriving simultaneously in different modalities as efficiently and rapidly as we do. Psycholinguists have investigated the integration of various sorts of linguistic information in the visual and auditory domain. In this thesis the structural and temporal aspects of the integration of letters and speech sounds are investigated. In Chapter 1, several questions were formulated concerning the relationship of grapheme and phoneme processing, three of which were the following.

First, can representational activation across modalities be demonstrated? Second, can such cross-modal activation, if it occurs, be of a facilitatory kind only, or inhibitory as well? Third, are any occurring cross-modal representational effects symmetric with respect to modality, i.e., is the activation effect from one modality to the other comparable in size and time-course to that in the other direction? As will be remembered, Chapter 3 only investigated the influence of grapheme context on phonemic processing, and not the opposite.

An explicit view on such cross-modal influences between sublexical orthographic and phonological representations in the context of visual word recognition is taken by the time-course model, briefly described in Chapter 3 (Seidenberg, 1985b, c; Seidenberg & McClelland, 1989). In this model, building on McClelland and Rumelhart's (1981) model of letter perception, reading words involves the computation of three types of representations in parallel: orthographic, phonological and semantic. If a visual word is processed, orthographic nodes corresponding to letter strings are activated, that subsequently activate associated phonological nodes (via a level of hidden nodes that do not concern us here). It is assumed that in principle phonological nodes feed activation back to orthographic nodes. Thus, the two types of nodes mutually activate or "coactivate" each other.

Though the implemented version of the model (Seidenberg & McClelland, 1989) uses distributed representations of grapheme and phoneme strings only, it would be in its spirit to assume that representations for congruent single graphemes and phonemes would also coactivate. For example, using the grapheme-phoneme correspondences of Dutch, activation of the grapheme node for the letter A should lead to a spreading of activation to the phoneme node for the speech sound /a:/ (as in the model of Sejnowski and Rosenberg, 1986). Since deactivating connections between between incongruent orthographic and phonological representations are absent, no cross-modal inhibition effects would be predicted at the representation level.

To test these predictions, I examined the relationship between the visual and auditory processing systems by means of a *bimodal vowel-detection task* involving letters and speech sounds. In the bimodal detection task subjects monitor two information sources in different modalities and give a speeded detection response to a pre-specified signal on either or both channels. An advantage of this task is that it leads to relatively fast reactions, thus making influences of slow attentional processes on the RT less likely. Moreover, unlike a cross-modal matching task (cf. Posner's experiments reported in Chapter 2), the detection task in principle allows the subject to react to bimodal stimuli on the basis of modality-specific representations; cross-modal contacts are not explicitly forced by the requirements of the task.

To set the stage for a discussion of my research, I will now describe this task in more depth, exploring how it is thought to be performed in the non-linguistic domain. Research using this task has centered upon the so-called Redundant Signals Effect (RSE; Kinchla, 1974), which refers to the finding that reaction times (RTs) to a redundant target stimulus (e.g., both a flash and a tone) are typically faster than those to either stimulus presented in isolation (e.g., to only a flash or a tone).

Two classes of models have been proposed to account for this effect:

separate activation models and coactivation models (Miller, 1982). According to *separate activation models* inputs on different channels are processed separately but in parallel. Both channels collect stimulus evidence (called "activation" by Miller, 1982) and as soon as a target is detected in either, a response is initiated. The RSE is thus seen by separate activation models as a result of the "race" between two temporally overlapping detection processes of randomly varying durations. It is basically considered as an effect of "statistical facilitation" (Raab, 1962), of which the size depends on the duration and the overlap of the distributions of the processes involved.

In *coactivation models* activation from different channels may be combined during processing to satisfy a single criterion for response initiation. Since activation is assumed to build up gradually over time until the criterion is reached, two channels *combining* their activation will, on average, lead to faster responses than only one source; or, in other words, to a RSE.

While separate activation models exclude coactivation as a contributing factor to the RSE, coactivation models do not exclude statistical facilitation (cf. Miller, 1982, p. 249; Miller, 1986, p. 332). Indeed, it is a common assumption in both types of models that processing a stimulus involves a number of steps, each of which may take a variable amount of time. This may easily lead to statistical facilitation in the RTs to redundant signals that are temporally close (cf. Ulrich & Giray, 1986).

While both separate activation models and coactivation models predict a RSE, they differ in how large they predict this effect to be. Miller (1982, 1986) has developed an analysis technique which can specify the limits of the facilitation predicted by separate activation models; if more facilitation occurs, this argues for the existence of coactivation, and the entire class of separate activation models can be rejected. However, if less or equal facilitation is found it is impossible on the basis of Miller's tests (described under Method of Analysis) to distinguish between separate or coactivation models.

Miller (1982) concluded that data obtained with detection tasks and letter search tasks were inconsistent with separate activation models and favored coactivation models, a conclusion that has not gone unchallenged (e.g., Van der Heijden, Schreuder, Maris, & Neerincx, 1984). According to Miller coactivation can occur at three different levels of processing: recognition, decision and response. Miller suggested that coactivation in his experiments had its primary locus at the decision stage.

By using signals that were unrelated (e.g., asterisks and tones) Miller excluded the recognition level as a possible locus for coactivation in his experiments. From a psycholinguistic point of view, however, cross-modal coactivation effects at a recognition level are of great interest, for example between orthographic and phonological word representations (e.g., Kirsner, Milech, & Standen, 1983; Monsell & Banich, 1982; Bradley & Forster, 1987), or, as in this thesis, at the sublexical level of grapheme and phoneme units (cf. Chapter 3).

On the basis of the psycholinguistic literature reported in Chapter 2 and that just referred to, it seems useful to distinguish two types of coactivation effects at a recognition level. Not only can the term "coactivation" be used to indicate an influence of two channels on a common representation (e.g., a modality-neutral unit for a grapheme or phoneme), but it may also refer to an influence between two connected units at the same level of representation (e.g., between a modality-specific grapheme and its corresponding phoneme). However, in both cases the response in an experimental task would depend on a representation that is "coactivated" by sensory information from two modalities.

With these views on coactivation in mind, I applied the bimodal detection task to investigate the psycholinguistic issues described above. In my experiments Dutch subjects detected visual and auditory targets such as the letters A and U and/or the speech sounds /a:/ and /u:/.[1] If cross-modal activation takes place, RTs to redundant trials with congruent letters and speech sounds (e.g., involving both the target letter A and the nominally identical target sound /a:/) should be faster than RTs predicted on the basis of statistical facilitation only. To show that such coactivation should be located at least in part at a representation level, the redundant condition just mentioned was compared to a redundant condition in which target letters and target speech sounds were incongruent (e.g., letter U and sound /a:/). Less coactivation was expected in the latter condition, since cross-modal activation at the representation level here should be absent (of course, other types of coactivation, for example at the decision level, could still occur).[2]

In order to exclude the possibility that differences between the two conditions mentioned would in part be due to cross-modal inhibition, a third redundant condition was introduced with a non-letter symbol (e.g., "*"), as the visual target stimulus. Since there is no phonological representation corresponding to this symbol and since such a symbol fulfills different functions than a letter or sound, it was assumed that any cross-modal inhibition at a representation level would be absent in this case.

To examine the symmetric or asymmetric nature of cross-modal influences, I varied the temporal relationship (SOA) between the presentation of visual and auditory stimuli. Assuming that a subject responds most often to the first presented target, RTs to a visual target followed by an auditory one should especially reflect phoneme-to-grapheme activation, RTs to an auditory target followed by a visual one should reflect grapheme-to-phoneme activation. If we now suppose, for example, that there is only a cross-modal influence of graphemes on phonemes, more coactivation should be present in the RTs when the first target is auditory than when it is visual.

4.1 Method of analysis

Since RTs to different single-channel stimuli will probably vary both within and between modalities due to stimulus characteristics (e.g., frequency, saliency, etc.) (Appelman & Mayzner, 1981), just comparing congruent (e.g., letter A and speech sound /a:/, abbreviated by Aa) and incongruent (e.g., letter U and sound /a:/, or Ua) redundant trials will not suffice. As has been argued in the introduction of this chapter, the amount of statistical facilitation to be expected for a redundant trial will vary with the amount of overlap and form between RT-distributions of the single channels involved, and the comparison between the different types of redundant conditions must take this fact into account.

Therefore, a comparison of the obtained RTs in the redundant conditions with RTs predicted on the basis of (maximal) statistical facilitation is indispensable to evaluate any results in this type of task. For this prediction, techniques developed by J. Miller and others make use of the empirically obtained single-channel conditions. The rationale behind those techniques will now be explained.

One type of separate activation model assumes independent channels. For such a model, processing in a redundant trial may be likened to a horse race. If two horses V and A race, the chance that the race is won by one of the two horses at time t is equal to the chance that horse V has finished at that time plus the chance that horse A has finished, minus the chance that both have finished; in mathematical terms:

$$(1) \quad P(RT_{va} \leq t) = P(RT_v \leq t) + P(RT_a \leq t) - P(RT_v \leq t \ \& \ RT_a \leq t)$$

Here, with independent channels, the last term is equal to the product $P(RT_v \leq t)P(RT_a \leq t)$. If reactions to redundant trials are considered the result of such a race between the visual and auditory channels, the above formula can be used to derive the predicted minimum distribution resulting from independent auditory and visual single-channel distributions.

To compare the information about the distributional characteristics of the RTs in the obtained and predicted conditions and to test separate activation models *in general* against coactivation models, Miller (1982) uses the fact that the last term in Equation 1 is always equal to or greater than 0, i.e. that

$$(2) \quad P(RT_v \leq t \ \& \ RT_a \leq t) \geq 0.$$

It follows, that with separate activation, for all values of t,

$$(3) \quad P(RT_{va} \leq t) \leq P(RT_v \leq t) + P(RT_a \leq t).$$

If this inequality is violated, *all* separate activation models (whether dependent or independent), have to be rejected: That is, when the (estimated) probability of occurrence of latencies smaller than some value t in the redundant condition exceeds the sum of the (estimated) probabilities in the two single-channel conditions. Coactivation models are consistent with violation of Inequality 3, because, with pooling of activation, the fastest responses to redundant signals can be faster than the fastest response to each channel alone (for further discussion of the method of analysis, see Appendix 1).

To evaluate separate activation models under signal conditions where one signal precedes the other by a certain time lag, Miller (1986, p. 332) has extended Inequality 3. According to such models, a response to redundant signals is caused by the first to finish of the two separate processes responding to each signal. If signal presentation is asynchronous, the two processes do not start at the same time and the finishing times must be adjusted to take that fact into account. If RT is measured from the onset of the first signal, the SOA between the two signals must be added to the latency of the responses to the second signal. Under such circumstances the following inequality should hold:

$$(4) \quad P(RT_{va} \leq t) \leq P(RT_v \leq (t - SOA_v)) + P(RT_a \leq (t - SOA_a)) \text{ for all t.}$$

In this Inequality 4, SOA_v and SOA_a denote the SOAs from the onset of the first signal to the onset of the visual and auditory signals, respectively. Either SOA_v or SOA_a will be zero in a particular redundant trial. When the visual signal precedes the auditory by an SOA_a, for example, SOA_v will be zero.

After testing for the existence of coactivation, the next step should be to compare the amount of coactivation in congruent (e.g., Aa) and incongruent (e.g., Ua) conditions. Under the assumption that the contribution of decision and motor level coactivation is comparable in both conditions[3], the extra contribution of coactivation at a representation level in the congruent condition should lead to more coactivation overall.

In the past, the significance of differences in amount of coactivation has sometimes been determined by running a by-subject ANOVA with Distribution (obtained curve vs. curve predicted by Inequality 3) as a factor (Miller, 1978). For the experimental situation described in this chapter, an interaction between the factors Congruence (congruent or incongruent conditions) and Distribution would then indicate that the amount of coactivation in the congruent and incongruent conditions was unequal. With this type of analysis, we would be testing the amount of coactivation with respect to a model that assumed maximal statistical facilitation for congruent and incongruent conditions (because it adds the visual and auditory probabilities as in Inequality 3). However, this type of extreme model does not seem very probable from a psychological point of view, since it would imply a high negative correlation between the two channels (fast processing in one channel would always occur with slow processing in the other, cf. Ulrich & Giray, 1986). (Further explanation of these issues is given in the Appendix 1).

Because of our ignorance concerning the correlation of the two channels, testing against a simple race model (assuming independent channels) seems to make more sense (cf. also Chapter 7, where it is shown how the response in this type of divided attention task can be seen as the result of a race between channels, modified by coactivation). In this chapter my approach will be to demonstrate first the existence of coactivation by means of Miller's method (using Inequalities 3 and 4), and to test subsequently the RT-differences between congruent and incongruent conditions after taking into account the predictions of the race model in the following way.[4]

Equation 1 was used to compute for a specific bimodal condition and subject the expected mean if the race model with independent channels would hold (using the obtained visual and auditory single-channel data as estimators for $P(RT_v \leq t)$ and $P(RT_a \leq t)$). This mean was subsequently subtracted from every RT for that condition and subject, and finally the thus corrected distributions were tested against each other. An advantage of this approach is that this test involves no distributions with special characteristics (such as the sum-curve which has a maximum cumulative probability of 2), and that the distributions involved, changed only by a constant, retain most of their properties.

4.2 Experiment 4

In the experiments to be reported, subjects were asked to respond as soon as possible if they detected a certain letter or symbol (e.g., A, U or &) and/or a certain speech sound (e.g. /a:/). Two kinds of comparisons were based on the obtained RT-data. First, RTs to bimodal stimuli consisting of both a visual and an auditory target were compared to RTs expected on the basis of independent separate activation of each. Second, RTs to bimodal stimuli where the visual and auditory targets were congruent (e.g., letter A and sound /a:/ in Dutch, abbreviated by "Aa") were compared with RTs to stimuli where they were incongruent (e.g., letter U and sound /a:/ or "Ua"), and with RTs to stimuli where one target was not a letter but a symbol (e.g., & and /a:/ or "&a").

4.2.1 Method

Subjects. Thirty-six undergraduates at Nijmegen University, all native speakers of Dutch, were paid to participate in the experiment.

Design. The experiment consisted of three sets of experimental trials, each of which involved a different combination of instruction and stimulus material (called Target Set from now on). In each Target Set three types of trials occurred: a visual stimulus was presented, an auditory stimulus was presented, or both a visual and an auditory stimulus were presented. In all Target Sets the letter E and sound /e:/ occurred as non-target stimuli, whereas the sound /a:/ was always the auditory target. Target Sets differed with respect to the visual target stimuli in the single-channel and redundant (bimodal) conditions. In Target Set Aa the visual target stimulus was the letter A, in Target Set Ua the letter U and in Target Set &a the symbol &. No trials combined target with non-target stimuli.

In bimodal trials the visual stimulus preceded the auditory one by 100 ms. This SOA (SOA=−100 ms)[5] was chosen because Miller (1986) found the largest effects of coactivation at SOAs around −100 ms.

Single-channel trials (auditory-only and visual-only) and bimodal signal trials were repeated 50 times. Thus, each Target Set had the following dimensions: 3 (type of trial) * 2 (target/non-target stimulus) * 50 (repetitions), giving 300 test stimuli. Furthermore, 20 practice trials were added, bringing the total number of trials in each Target Set to 320.

Stimuli. The auditory stimuli were recorded on tape by a female native speaker of Dutch in a sound-proof room. The length of the vowel /a:/ was 280 ms, while /e:/ had a duration of 350 ms. The stimuli were digitized on a VAX 11/750 computer with a sampling rate of 20 kHz. For each Target Set a randomized sequence of targets was placed on one channel of a tape. The output of the computer was low-pass filtered with a cutoff frequency of 10 kHz. During the experiment the auditory stimuli were presented binaurally over headphones. On the second channel of the tape a pulse, inaudible to the subjects, was placed, that triggered both the timer for the recording of the RTs and the presentation of the visual stimulus.

The visual stimuli were white Roman capitals, 6 mm in height, presented on a MATROX-screen with a dark background. The monitor was placed at a distance of 60 cm from the subject, in order to provide projection within the foveal field of the eye. The visual stimuli were presented for 60 ms. Presentation of the visual stimuli and recording of the RTs were controlled by a PDP-11/23 computer.

Procedure. Subjects were tested on two successive days. On the first day two experimental Target Sets were run, on the second day the third. The order of Target Sets was counterbalanced over subjects.

Before running a Target Set, subjects read a written instruction. The instruction was repeated orally at the beginning of the experiment. Subjects were told to lightly rest the index finger of their preferred hand on the response button in front of them and to push this button as fast as possible whenever they saw or heard a specific visual or auditory stimulus, or a combination of these. At the beginning of Target Set Aa they were told to react when they saw the letter A, when they heard the sound /a:/ or when both the letter A and the sound /a:/ appeared. Before Target Set Ua they were instructed to react whenever they saw the letter U and/or heard the sound /a:/; before Target Set &a when the symbol & and/or the sound /a:/ occurred. Each time they were also told not to respond to visual and/or auditory presentations of "E". The task was therefore a go/no-go one.

Each trial started with a 1000 Hz. warning signal of 150 ms duration. In the visual-only condition, this warning signal was followed after 400 ms of silence by the visual signal. In the auditory-only condition, the period of silence between warning signal and auditory stimulus was 500 ms. The redundant trials combined the two presentations. Two seconds after presentation of the last signal a new trial was initiated.

Each Target Set consisted of 20 practice trials followed by a block of 300 test trials. After the practice trials there was a short pause in which subjects, if necessary, could ask for clarifications. On the first day two such Target Sets were presented in approximately one hour. Between the two Sets there was a two-minute break. On the second day one Target Set was run in about half an hour.

4.2.2 Results

Mean RTs (measured from the onset of the first presented target stimulus) were computed for each subject and each experimental condition in each Target Set. Latencies greater than 750 ms or smaller than 150 ms were treated as errors. One subject with more than 15% errors was excluded from later analyses. The subsequent total percentage of missing and extreme values was 1.4%. Errors were substituted by mean RTs in each subcondition for each subject. The percentage of "false alarms", i.e. the reaction to the no-go trials, was 1.9%. Table 4.1 shows the main results.

	Condition		
Target Set	redundant	visual	auditory
Aa	315	334	349
Ua	378	393	386
&a	362	372	376

Table 4.1. Mean RTs (in ms) in the redundant auditory and visual conditions, for three Target Sets: A-/a:/ or Aa; U-/a:/ or Ua; &-/a:/ or &a.

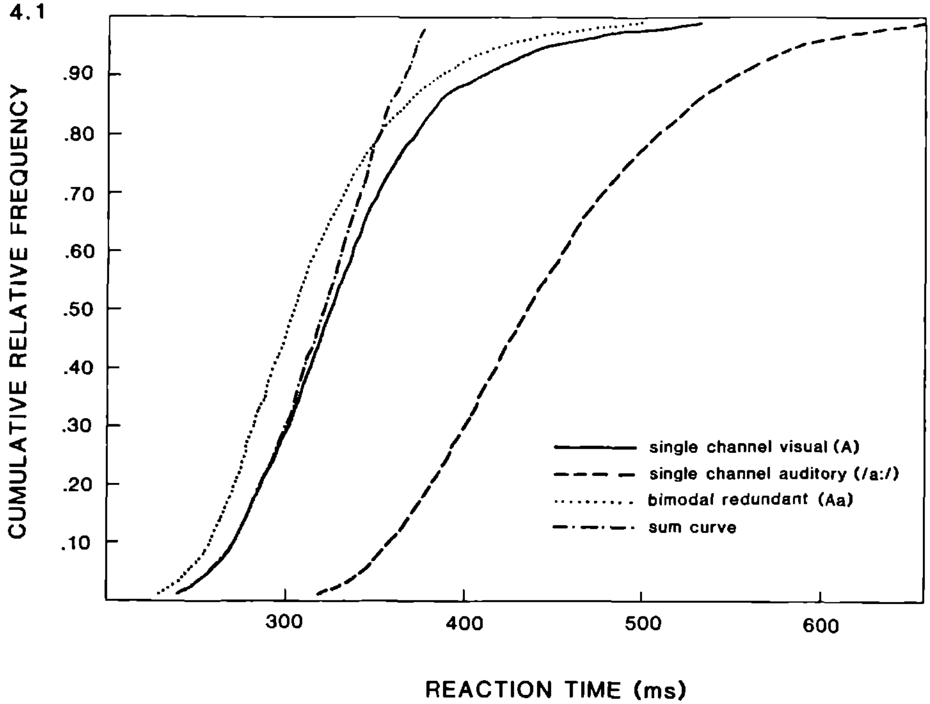
In order to check for effects of coactivation I computed the average Cumulative Distribution Functions (CDFs) for the visual-only, auditory-only and redundant conditions in each Target Set, by averaging across subjects (cf. Ratcliff, 1979). To obtain these CDFs (depicted in Figures 4.1, 4.2 and 4.3), RTs in each condition for a given subject were rank ordered. If any of the 50 RTs were smaller than 150 or larger than 750 ms, a full distribution was generated by means of a damped cubic SPLINE-function (cf. de Boor, 1978). Each of the 50 ordered RTs estimates the RTs at the 1st, 3rd, and 5th-99th percentile of the true CDF for a given subject. Composite CDFs were then formed by averaging, across subjects, all of the RTs for a given percentile (cf. Miller, 1982).

These figures also show the redundant signal CDFs as compared with the sum of the single-channel CDFs, displaying the comparison represented in Inequality 3. The inequality was violated throughout the range from the 1st to the 73th percentiles of RT for Target Set Aa, from the 1st to the 19th percentiles of RT for Target Set Ua and from the 1st to the 23th percentiles of RT for Target Set &a, as demonstrated by paired t-tests ($p<.05$) between redundant signal and sum-curve distributions across subjects at each of the 50 percentile points in a Target Set.

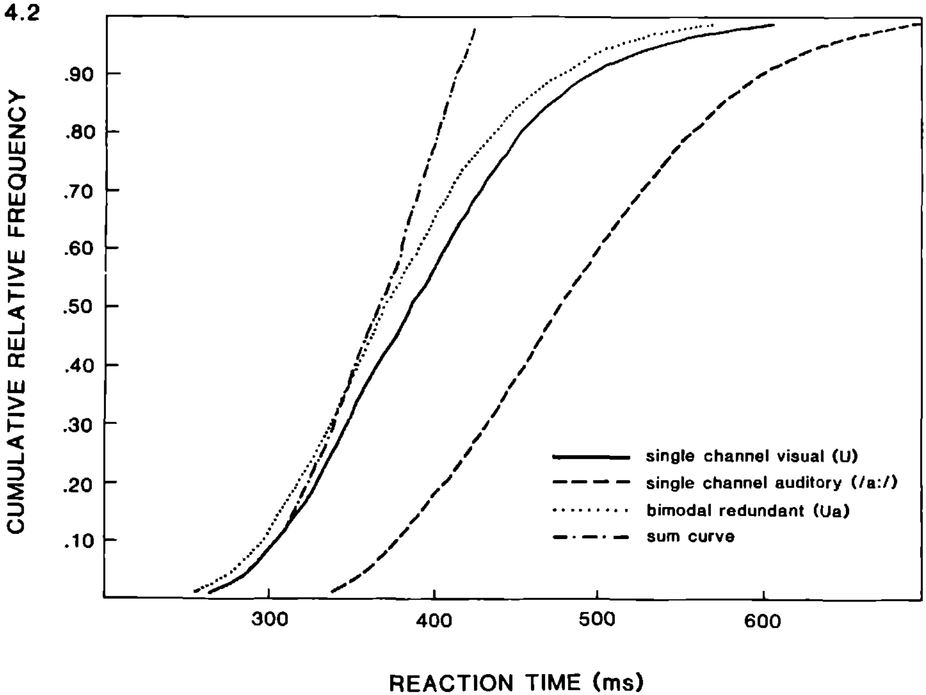
After this test for coactivation, I compared the different Target Sets with respect to their amount of deviation from an independent race model. For each subject and each redundant condition, the predicted minimum distribution was computed using the single-channel data (following Inequality 1 above; see Appendix 1 for more information). After adding 100 ms to the raw RTs of the auditory-only trials (SOA_a), the correct RTs were ordered in ms steps for the analysis to obtain the highest resolution in the predictions. Only RTs between 150 and 750 ms were included in the analysis. The mean bimodal RTs computed from the predicted minimum distribution given Inequality 3 are shown in Table 4.2, together with the means obtained and the correlation between the means obtained and predicted.

For each subject, the predicted mean reaction time for a redundant condition was subtracted from each raw reaction time obtained in that condition. After this correction for statistical facilitation, the resulting times were used as estimates of the amount of coactivation. To test for differences in the amount of coactivation among Target Sets, an ANOVA was subsequently carried out on the thus corrected redundant conditions. This analysis showed a main effect of Target Set [$F(2,68)=15.59$, $p<.001$].

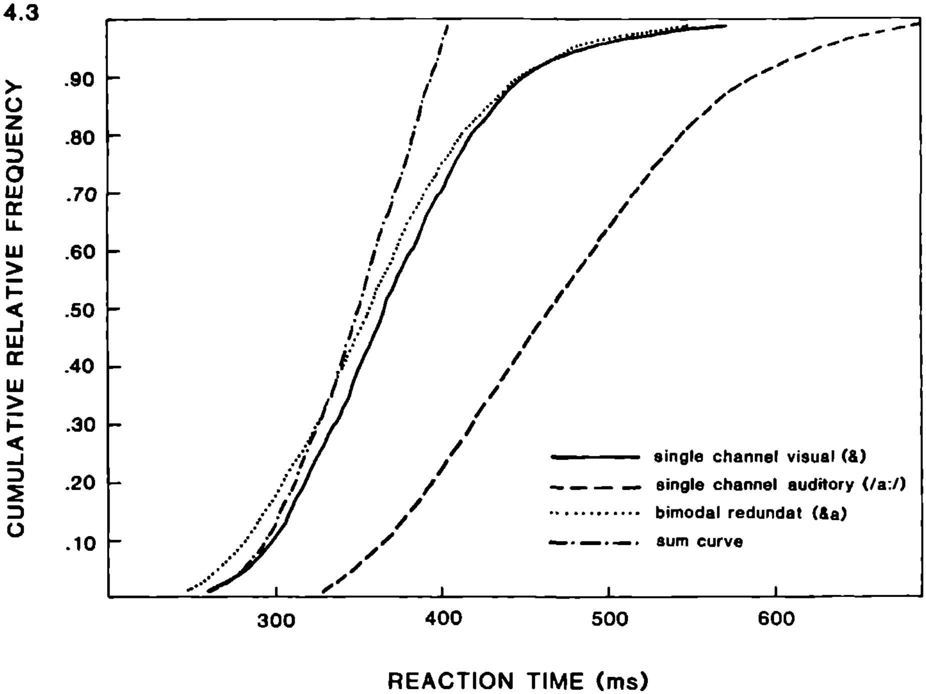
4.1


CUMULATIVE RELATIVE FREQUENCY
.90
.80
.70
.60
.50
.40
.30
.20
.10
300
400
500
600
REACTION TIME (ms)
single channel visual (A)
single channel auditory (/a:/)
bimodal redundant (Aa)
sum curve

4.2


CUMULATIVE RELATIVE FREQUENCY
.90
.80
.70
.60
.50
.40
.30
.20
.10
300
400
500
600
REACTION TIME (ms)
single channel visual (U)
single channel auditory (/a:/)
bimodal redundant (Ua)
sum curve

4.3



Figures 4.1, 4.2, and 4.3. Cumulative distribution functions (CDFs) for redundant, visual-only and auditory-only conditions and the sum of the single-channel CDFs for Target Sets Aa, Ua, and &a.

Paired planned comparisons showed that this effect was due to a difference between Target Sets Aa and Ua ($t(34)=4.78$, $p<.001$), and Target Sets Aa and &a ($t(34)=5.15$, $p<.001$) but not to a difference between Target Sets Ua and &a ($t(34)=.28$, ns).

Finally, correlations were computed between the obtained RTs and the RTs predicted on the basis of the race model. As can be seen in Table 4.2, these correlations were all highly significant.

Target Set	predicted	obtained	t(34)	p-value	corr
Aa	326	315	−6.47	<.001	.93
Ua	375	378	.99	ns	.93
&a	358	362	1.10	ns	.90

Table 4.2. Mean RTs (in ms) for redundant conditions as predicted by an independent separate activation model and as obtained, the significance of their differences, and correlations between predicted and obtained means for three Target Sets: A-/a:/ or Aa; U-/a:/ or Ua; &-/a:/ or &a. All correlations significant at $p < .001$.

4.2.3 Discussion

The results of the experiment confirm the existence of cross-modal activation between phoneme and grapheme representations. First, the RTs in Target Set Aa were significantly shorter than those predicted by a model that assumes an independent race between the visual and auditory target signals (cf. Equation 1), while those in Target Sets Ua and &a did not differ from the predictions. The analyses of the experimental results further showed that the differences between obtained and predicted RTs were significantly larger in Target Set Aa than in the other Sets. Finally, application of Miller's test to the obtained data indicated the presence of coactivation.

More specifically, the results suggest activation of graphemes by congruent phonemes. From the single-channel RTs it can be inferred that processing the letter took about as long as processing the auditory vowel. Since the visual signal preceded the auditory by 100 ms, it therefore seems likely that in the redundant conditions subjects reacted predominantly to the letter. Thus, given the facilitation effects in Target Set Aa compared to the other Sets, the graphemic representations on which the redundant responses were based in this Target Set were probably influenced by their phonemic counterparts.

This conclusion is of considerable interest, since the existence of fast phonemic influences on graphemes has not been investigated in much depth before (a few studies are available that examined such influences in the context of word recognition: Seidenberg & Tanenhaus, 1979; Tanenhaus, Flanigan, & Seidenberg, 1980; Donnenwerth-Nolan, Tanenhaus & Seidenberg, 1981). One reason for this may be that, while a certain dependence of visual word recognition on auditory processing systems can be motivated by phylogenetic and ontogenetic arguments (cf. Scinto, 1986), the opposite is harder to maintain (but see Ehri, 1985, for an alternative view involving a mutual dependence).

Going further, I want to argue that the cross-modal activation probably arose before the identification of the visual representation was completed. Since in both Target Set Aa and Ua the visual target was associated with a go-response, the possibility is excluded that different amounts of coactivation at the decision or response levels led to the different RT-patterns observed. Therefore, the facilitation effect should have arisen at the representational level before the identification of the target was completed and the decision to respond could be made.

The RT-patterns in the Ua Target Set were very similar to those in the &a Target Set. This could be interpreted as evidence for a lack of inhibition between phonemes and incongruent graphemes (or the phonemes activated by those graphemes) in Target Set Ua, under the assumption that for a non-letter symbol like "&" inhibitory activation from phonemes is absent.

However, the absence of inhibition effects could be ascribed to the relatively short presentation of the visual stimulus (60 ms). Perhaps cross-modal inhibition effects need a longer stimulus duration to build up. Furthermore, the choice of an SOA of −100 ms may have led to the absence of inhibition effects in the Ua-Target Set: If cross-modal inhibition effects take more time to develop or are smaller than facilitation effects, perhaps there was not sufficient time for a significant effect to arise under this SOA.

My choice of this SOA was motivated by Miller's (1986) finding of the largest effects of coactivation for letters and tones in this range of SOA. However, re-inspection of his data showed that RTs in his single-channel conditions differed up to 100 ms for the visual and auditory modality (p. 335). In my experiment overall RT-differences to letters and phonemes were much smaller (maximally 15 ms); therefore, a simultaneous presentation of both signals might have provided clearer facilitation and inhibition effects.

The reader will have noticed that no ANOVAs on the raw data were reported. As argued in the Introduction, such ANOVAs on the raw data neglect the varying relationship between the visual and auditory single-channel conditions, and the corresponding redundant condition. The high correlations between the RTs obtained in the redundant conditions and the RTs predicted on the basis of a race between the component visual and auditory stimuli stress the relevance of the correction method I applied. Speed of processing in a redundant condition to a large extent seems to depend on and vary with the processing speeds of the contributing visual and auditory targets (as reflected by the RTs to the targets in isolation). For a correct estimation of the representational effects in the RTs of congruent as compared to incongruent conditions, the processing characteristics of the various visual and auditory targets must be taken into account.

I therefore felt justified to continue using the correction method in the next experiment, which was intended to replicate the results of Experiment 4 for an SOA of −100 ms (visual signal leading by 100 ms), and expand it by adding an SOA of 0 ms (simultaneous presentation of letter and sound). The component single-channel distributions will probably have a larger overlap at an SOA of 0 ms than at an SOA of −100 ms, which will lead to more statistical facilitation and therefore to faster reaction times in redundant conditions. However, since our correction method takes into account differences in statistical facilitation between conditions and SOAs, this is of no concern to us. Of more interest is whether coactivation effects at representational and non-representational (e.g., at decision or response) levels increase from SOA1=−100 ms to SOA2=0 ms, due to the larger overlap of the distributions involved at SOA2. Finally, in this experiment the visual target stimulus was presented for a duration comparable to the auditory one, and the "&" symbol was replaced by a star ("*"). This last change was made to exclude a possible linguistic interpretation of the symbol, since in Dutch "&" is pronounced as "/en/", the letter-name for "N".

4.3 Experiment 5

4.3.1 Method

Subjects. Thirty-one undergraduates at Nijmegen University, all native speakers of Dutch, were paid to participate in the experiment.

Design. The design of this experiment was very similar to that of Experiment 4. However, the redundant trials were presented under two SOAs. For SOA1 the visual stimulus appeared 500 ms after the warning signal, 100 ms before the auditory stimulus was started. For SOA2 the visual and auditory stimulus had a simultaneous onset, 600 ms after the warning signal. In the visual-only condition, the visual stimulus was presented either 500 or 600 ms after the warning signal, depending on SOA. The auditory stimulus always appeared 600 ms after the warning signal; to balance the number of visual and auditory trials, an equal number of auditory-only trials were allocated to SOA1- and SOA2-conditions. The warning signal at the onset of a trial in this experiment had a duration of 200 ms. All visual stimuli were displayed for 280 ms, the duration of the auditory /a:/-target stimulus.

Also, the number of repetitions of each type of trial was reduced to 40. Each of the three Target Sets (Aa, Ua and *a) was run in a separate session and had the following dimensions: 3 (type of trial) * 2 (target/non-target stimulus) * 2 (SOA) * 40 (repetitions) = 480 test stimuli. An additional 48 practice trials were constructed, leading to a total number of 528 trials in a Target Set.

Procedure. Subjects participated in three sessions on successive days. The order of Target Sets in these sessions was counterbalanced over subjects. Each experimental session lasted for about 45 minutes, and had a short break after about 25 minutes.

4.3.2 Results

Mean RTs (measured from the onset of the first presented target stimulus) were computed for each subject and each experimental condition. Latencies greater than 750 ms or smaller than 150 ms were treated as errors. The total percentage of missing and extreme values was 1.1%. Errors were substituted by mean RTs in each subcondition for each subject. The percentage of "false alarms", i.e. reactions to the no-go trials, was 2.1%. Table 4.3 shows the main results, analyzed for each SOA separately.

To test for coactivation in the different redundant conditions under each SOA, I applied Miller's technique. However, before applying it to SOA2 (simultaneous presentation), I followed Miller's (1982) suggestion to compute the average across subjects for the faster of the two single-channel conditions in a Target Set. The three resulting values were 328 (Aa), 341 (Ua) and 343 (*a). Testing against the obtained values in the redundant conditions of respectively 290, 317, and 310, significant differences were found in all cases (Aa: $t(30)=8.46$, $p<.001$; Ua: $t(30)=6.43$, $p<.001$; *a: $t(30)=9.61$, $p<.001$). This result is important because it indicates that Redundant Signals Effects obtained in each Target Set will not be an artefact of averaging across some subjects who detected the visual signal faster, and other subjects who detected the auditory signal faster (Miller, 1982, p. 255). (Such a detection difference would be quite unlikely for SOA1, where the visual signal appeared 100 ms before the auditory; therefore no such test was performed for this SOA).

SOA1=−100 (letter leads by 100 ms)

Target Set	redundant	visual	auditory
Aa	328	345	345
Ua	360	376	348
*a	351	368	351

SOA2=0 (simultaneous onset of letter and speech sound)

Target Set	redundant	visual	auditory
Aa	290	340	345
Ua	317	373	351
*a	310	358	360

Table 4.3. Mean RTs (in ms) in the redundant, auditory-only and visual-only conditions, for SOA's of −100 ms (visual leading) and 0 ms in three Target Sets Aa, Ua and ***a**. RTs to auditory-only trials are to identical tokens presented 600 ms after the ending of the warning signal, RTs to visual-only trials are to identical tokens presented at 500 or 600 ms after ending of the warning signal.

The average CDFs were computed for the visual-only, auditory-only and redundant conditions in each Target Set and under each SOA, computed by averaging across subjects. To obtain these CDFs, RTs in each condition for a given subject were rank ordered (for SOA1 after adding 100 ms to the RTs in the auditory-only condition). If any of the 40 RTs were smaller than 150 or larger than 750 ms, a full distribution was generated by means of a damped cubic SPLINE-function. Each of the 40 ordered RTs estimates the RTs at the 1.25th. 3.75th, and 6.25th-98.75th percentile of the true CDF for a given subject. Composite CDFs were then formed by averaging, across subjects, all of the RTs for a given percentile (cf. Miller, 1982). The redundant signal CDFs were compared with the sum of the single-channel CDFs, represented in Inequality 3 and 4. For SOA1=−100 ms, the inequality was violated throughout the range from the 1.25th to the 61.25th percentiles of RT for Target Set Aa with the exception of percentile 3.75 ($t(30)=.19$), and from the 6.25th to the 46.25th percentiles of RT for Target Set *a, as demonstrated by paired t-tests between redundant signal and sum-curve distributions across subjects at each of the 50 percentile points in a session. It was never significantly violated under this SOA for Target Set Ua. For SOA2=0 ms, the inequalities were violated from the 1.25th to the 61.25th percentile for Target Set Aa, from the 1.25th to the 26.25th percentile for Target Set Ua, and from the 7.75th to the 33.75th percentile for Target Set *a.

After this demonstration of the existence of coactivation, I wanted to compare the redundant conditions in the different Target Sets, taking differences resulting from statistical facilitation into account. For each subject and each redundant condition, first the predicted minimum distribution was computed, using Inequality 1. After adding 100 ms to the RTs of the auditory-only condition under SOA1=−100 ms, the correct RTs were ordered in ms steps for the analysis to obtain the highest resolution in the predictions. Only RTs between 150 and 750 ms were included in the analysis. The mean predicted bimodal RTs computed from the predicted minimum distribution are given in Table 4.4 for SOA1=−100 ms and for SOA2=0 ms, together with the means obtained, the significance of their difference and the correlation between the means obtained and predicted.

For each subject, the mean obtained reaction time for a certain redundant condition under a specific SOA was subtracted from each reaction time predicted for that condition and with that SOA. Subsequently, an ANOVA was conducted using the adapted redundant conditions, showing a main effect of Target Set [$F(2,60)=8.27$, $p<.001$], but not of SOA and no interaction between Target Set and SOA ($F<1$ in both cases). Thus, SOA1=−100 ms and SOA2=0 ms did not differ significantly in the size of facilitation effects with respect to an independent separate activation model.

SOA1=−100 (letter leads by 100 ms)

Target Set	predicted	obtained	t(30)	p-value	corr
Aa	335	328	−2.57	<.05	.96
Ua	358	360	.60	ns	.94
*a	354	351	−1.25	ns	.96

SOA2=0 (simultaneous onset of letter and speech sound)

Target Set	predicted	obtained	t(30)	p-value	corr
Aa	301	290	−2.82	<.01	.92
Ua	315	317	.35	ns	.93
*a	312	310	−.83	ns	.93

Table 4.4. Mean RTs (in ms) for redundant conditions as predicted by an independent separate activation model, and as obtained, the differences between the predicted and obtained means and their correlations for three experimental Target Sets Aa, Ua and *a, at SOA1=−100 ms and SOA2=0 ms. All correlations significant at $p < .001$.

Disregarding SOA, paired planned comparisons, performed on the differences between the corrected redundant conditions, showed significant differences between Target Sets Aa and Ua ($t(30)=-4.14$, $p<.001$); and between Aa and *a ($t(30)=-2.17$, $p<.05$). The difference between Ua and *a was marginally significant ($t(30)=1.83$, $p=.08$).

Finally, ANOVAs on the raw data tested for differences among the single-channel conditions across SOAs and Target Sets. For the auditory single-channel conditions, no significant main effect of Target Set [$F(2,60)<1$] was found, indicating that the auditory single-channel conditions were comparable over Target Sets. No significant difference was found either between auditory-only trials under SOA1 and SOA2 [$F(1,30)=2.73$, $p>.10$], agreeing with the fact that all trials consisted of identical tokens. The interaction between the Target Set and SOA was not significant either [$F(2,60)=2.40$, $p=.10$). For the visual single-channel conditions, different results were obtained. Significant differences were found among Target Sets [$F(2,60)=13.79, p<.001$] and between SOA1 and SOA2 [$F(1,30)=12.80$, $p<.001$), for which the moment of stimulus onset (at 500 or 600 ms after the warning signal) differed by 100 ms. The interaction between Target Set and SOA was not significant [$F(2,60)=1.35$, $p>.10$).

4.3.3 Discussion

The results of Experiment 5 replicated those of Experiment 4: RTs in Target Set Aa were faster than those in both Ua and *a Target Sets after correcting for the characteristics of the contributing single-channel distributions. Also, the obtained curves for Target Set Aa showed a violation of the sum-curve distribution over a much longer range than did Target Sets Ua and *a. Following the same reasoning as before, these results provide evidence for cross-modal activation between phoneme and grapheme representations.

The inclusion of two SOAs in Experiment 5 yields some information concerning the time-course of this activation process. The RT-difference between the means obtained and those predicted on the basis of statistical facilitation in Target Set Aa is not significantly larger under SOA2 (simultaneous presentation) than under SOA1 (visual leads by 100 ms). This suggests that the cross-modal activation at the representation level was as strong under SOA1 as under SOA2. Taking into account that motor processes taking place after perceptual and decision processes take in the order of 90 ms or more (cf. Carlton, 1981; Gottlieb & Agarwal, 1973; Gaillard & Perdok, 1980) and that there was an SOA of 100 ms, this points to a contact between the visual and auditory representation before about 330-90-100 = 140 ms, supporting my conviction that I am measuring early, representational effects here.

Consistent with the results of Experiment 4, again high correlations were found between the RTs obtained and predicted for the redundant conditions (all above .90). The importance of statistical facilitation suggested by this finding is strengthened further in that the strong increase in the RSE going from SOA1 to SOA2 seems primarily due to an increase in such facilitation (caused by the larger distributional overlap under SOA2): The differences in RTs obtained and RTs predicted on the basis of a race between two independent channels remain as small as before.[6]

While both Experiment 4 and 5 showed significant results in favor of the hypothesis of cross-modal activation, the effects observed were rather small in size. It is therefore important to exclude alternative explanations for those results. However, I have not been able to come up with any consistent alternative explanation for the pattern of results in Experiment 4 and 5. For example, one type of explanation could be construed in terms of differences in the visual discriminability of the various target and non-target stimuli: Perhaps the visual target A is less similar to the non-target E than target U is, thus leading to faster RTs for visual single-channel and redundant conditions in the Aa Target Set than in the Ua Target Set. However, since my prediction method takes into account differences in RTs between the single-channel conditions, this suggestion would still not explain why the facilitation effect with respect to the race prediction would be larger in the Aa Target Set than in the other Sets. Furthermore, as a test for the existence of such differences in target discriminability I performed a control experiment that completely replicated Experiment 4, except that only visual target and non-target stimuli were included. No RT-differences were obtained between blocks of trials with A and U target stimuli, both mixed with E non-target stimuli. This strongly suggests that the two target letters did not differ in visual discriminability from the non-targets.

Perhaps different explanations could be based on the assumption of other types of processing differences among the blocked Target Sets, which the over-all RT-differences between these Sets might reflect. To exclude any potential problems caused by differences in single-channel baseline conditions between the Target Sets, I decided to change from a blocked to a mixed presentation of stimuli in a third experiment. Indeed, if the effects found in Experiments 4 and 5 were genuine, automatic representational effects, they should be robust over strong changes in design.

Furthermore, Experiment 6 also included bimodal stimuli in which a target or non-target in one modality was paired to a "neutral" stimulus in the other. A neutral stimulus was not a target stimulus, and occurred equally often in go and no-go trials. As neutral stimuli were used the letter I, speech sound /i/, a star or white noise. The inclusion of this bimodal baseline-condition served two important goals. First, it could be used for a comparison to the raw data from the congruent and incongruent redundant conditions. Second, RT-differences between linguistic and non-linguistic neutral stimuli should yield some information concerning the depth of processing of the non-target stimulus and the presence of attention shifts from one modality to the other under different SOAs.

Finally, Experiment 6 added an SOA-condition in which the onset of the auditory stimulus preceded that of the visual stimulus by 100 ms.

Therefore, conditions occurred in which the visual stimulus preceded the auditory one, was presented simultaneous with it, or followed it. This made the experiment temporally symmetric with respect to the visual and auditory modality. I wanted to investigate whether this temporally more balanced design would lead to RT-differences in the SOAs used before, due to differences in the subjects' division of attention over the two modalities.

Including conditions in which the auditory stimulus precedes the visual one has interesting theoretical implications as well. I argued before that reactions to a visual target followed by an auditory congruent target 100 ms later predominantly reflects auditory-to-visual activation. Similarly, an auditory target followed by a 100 ms delayed congruent visual target should reflect visual-to-auditory activation. A comparison of the results of these two SOA-manipulations could thus potentially indicate a directional asymmetry in the size or speed of cross-modal activation (e.g., the influence from auditory-to-visual might be less strong than from visual-to-auditory). Such evidence concerning the mutual dependence of visual and auditory sublexical processing systems could be used in the construction of models simulating word recognition processes (cf. the inclusion of a phonological route in the time-course model for visual word recognition by Seidenberg and coworkers; and see Chapter 7).

The next experiment incorporated the various conditions just proposed. It included bimodal redundant stimuli (e.g., visual A combined with auditory /a:/, abbreviated by Aa), bimodal non-redundant stimuli (e.g., Ia) and single-channel stimuli (e.g., A alone) presented under three different temporal relationships (SOAs of −100, 0 and 100 ms). Also, all conditions were included in one completely mixed experiment, in order to reduce variances due to session and learning effects. Finally, as exemplars of congruent redundant conditions both Aa- and Uu-trials were included; Au- and Ua-trials made up the incongruent redundant conditions.

4.4 Experiment 6

4.4.1 Method

Subjects. Thirty-one undergraduates at Nijmegen University, all native speakers of Dutch, were paid to participate in the experiment.

Design. The experiment was run in three sessions. In all sessions 16 different stimulus presentation conditions occurred, as shown in Table

4.5. There were four bimodal redundant conditions, eight bimodal non-redundant conditions, and four single-channel conditions.
In the redundant and non-redundant conditions three SOAs were used, with the visual stimulus being presented 100 ms before, with the same onset time as, or 100 ms after the auditory stimulus.

Redundant conditions were repeated 20 times under all SOAs. Single-channel conditions were repeated 20 times. This leads to a total number of test stimuli in the experiment of: 12 (conditions) * 2 (target/non-target stimulus) * 3 (SOA) * 20 (repetitions) + 4 (single-channel conditions) * 2 (target/non-target stimulus) * 20 (repetitions)= 1440 + 160 = 1600 stimuli. Furthermore, 40 practice trials for each session were constructed so that the total number of stimuli presented in the experiment as a whole amounted to 1720. The number of trials in a session was therefore 573 (or 574).

As before, the auditory stimuli were recorded on tape by a female native speaker of Dutch in a sound-proof room. Naturally sounding stimuli of about equal length (320 ms) were chosen for use in the experiment. All visual stimuli were displayed for 320 ms, the average duration of the auditory stimuli.

	TARGETS		**NEUTRAL**			**NO-GO**	
AUDITORY:	/a:/	/u:/	/i/	NOISE	–	/o:/	/e:/
VISUAL							
TARGETS							
A	Aa	Au	Ai	An	A-		
U	Ua	Uu	Ui	Un	U-		
NEUTRAL							
I	Ia	Iu				Io	Ie
*	*a	*u				*o	*e
–	-a	-u				-o	-e
NO-GO							
E			Ei	En	E-	Eo	Ee
O			Oi	On	O-	Oo	Oe

Table 4.5. Stimulus conditions in Experiment 6, ordered by visual stimulus (column, first symbol) and auditory stimulus (row, second symbol). n stands for "NOISE", – for "no signal".

Procedure. The experiment was run over three successive days. The order of sessions was counterbalanced over subjects.

In this experiment, subjects were instructed to push the response button as fast as possible whenever they saw and/or heard the letter or sound "A" and/or the letter or sound "U" (Dutch). They were told not react if only other letters or sounds were presented.

Each trial started with a 1000 Hz. warning signal of 200 ms duration. In the visual-only and the auditory-only trials, this warning signal was followed after 600 ms by the target stimulus. In the redundant trials, the visual stimulus followed the warning signal after 500, 600 or 700 ms of silence (depending on SOA); the auditory stimulus was always presented after 600 ms. Two seconds after presentation of the auditory signal a new trial was initiated.

Each session consisted of 40 practice trials followed by a block of 533 (or 534) test trials. After the practice set there was a short pause in which there was an opportunity for asking questions. Each session lasted for about 45 minutes, with a three-minute break after about 25 minutes.

4.4.2 Results

Mean RTs (measured from the onset of the first presented target stimulus) were computed for each subject and experimental condition. Latencies greater than 750 ms or smaller than 150 ms were treated as errors. The total percentage of missing and extreme values was 1.3%. Errors were substituted by mean RTs in each subcondition for each subject. The percentage of "false alarms", i.e. reactions to the no-go trials was 1.6%. Figure 4.4 shows the main results for the redundant and single-channel conditions. For the exact RT-values see the associated Table in Appendix 2.

I will first analyze the results for the bimodal redundant conditions; afterwards separate analyses will be presented for the non-redundant conditions.

As before, I set out to test for coactivation. Before applying Miller's technique to SOA2=0 ms, the average across subjects for the faster of the two single-channel conditions in a session was computed. The four resulting values were 386 (Aa), 402 (Au), 389 (Ua) and 416 (Uu). Testing against the obtained values in the redundant conditions of respectively 334, 379, 364 and 369, significant differences were found in all cases (Aa: $t(30)=13.98$, $p<.001$; Au: $t(30)=5.52$, $p<.001$; Ua: $t(30)=4.94$, $p<.001$);

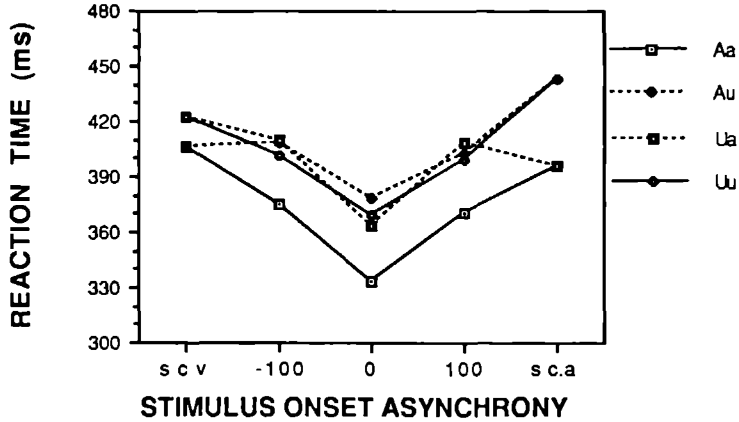


Figure 4.4. Mean RTs (in ms) in the redundant and single-channel conditions when the visual stimulus preceded the auditory stimulus (SOA1=−100 ms), accompanied it (SOA2=0 ms), or followed it (SOA3=100 ms). RTs were measured from the first presented target stimulus. S.c.v. stands for "single-channel visual", s.c.a. for "single-channel auditory".

Uu: $t(30)=11.10$, $p<.001$). This indicates that the Redundant Signals Effect obtained in each session is not an artefact of averaging across some subjects who detected the visual signal faster, and other subjects who detected the auditory signal faster (cf. Exp. 5).

The average CDFs for the visual-only, auditory-only and redundant conditions in each session and under each SOA were computed by averaging across subjects. To obtain these CDFs, RTs in each condition for a given subject were rank ordered (for SOA1 and SOA3 after adding 100 ms to the RTs in the single-channel condition of the second signal). If any of the 20 RTs were missing, a full distribution was generated by means of a damped cubic SPLINE-function. Each of the 20 ordered RTs estimates the RTs at the 2.5th, 7.5th, and 12.5th-97.5th percentile of the true CDF for a given subject. Composite CDFs were then formed by averaging, across subjects, all of the RTs for a given percentile (cf. Miller, 1982).

The redundant signal CDFs were compared with the sum of the single-channel CDFs, represented by Inequality 3 and 4. For SOA1=−100 ms, the inequality was violated throughout the range from the 2.5th to the 37.5th percentiles of RT for the Aa-condition and in percentile 27.5 and 32.5 for the Uu-condition, as demonstrated by paired t-tests between redundant signal and sum-curve distributions across subjects at each of the 50 percentile points in a session. It was never significantly violated under this SOA for the Ua- and Au-conditions. For SOA2=0 ms, the inequality was violated from the 2.5th to the 47.5th percentile for Aa, from the 2.5th to the 37.5th percentiles of RT for Uu, from the 2.5th to the 17.5th percentiles of RT for Au, and from the 2.5th to the 17.5th percentiles of RT for Uu.

For SOA3=100 ms, some puzzling results were obtained. The inequality was never significantly violated for Aa, but it was significantly violated for Au from the 2.5th percentile to the 42.5th percentile, for Ua for percentile 7.5 and 12.5, and for Uu from the 7.5th through 17.5th percentile, staying marginally significant over a longer range (e.g., at percentile 37.5: $p=.08$).

After this test for coactivation, I compared the congruent and incongruent redundant conditions in the following way, in order to take differences resulting from statistical facilitation into account. For each subject and each condition, the predicted minimum distribution was computed, using the single-channel conditions according to Inequality 1.

After addition of 100 ms to the single-channel RTs of the later signal for SOA1 and SOA3, the correct RTs were ordered in ms steps for the analysis to obtain the highest resolution in the predictions. Only RTs between 150 and 750 ms were included in the analysis. The mean predicted bimodal RTs computed from the predicted minimum distributions given Inequality 1 are given in Table 4.6 for all SOAs, together with the means obtained, the significance of their difference and the correlation between the means obtained and predicted.

For each subject, the mean predicted reaction time for a certain redundant condition under a specific SOA was subtracted from each reaction time obtained in that condition and with that SOA. Figure 4.5 depicts the adapted RT-patterns over SOA for the four redundant conditions (now indicated as Aa', Au', Ua' and Uu'). Subsequently, an ANOVA was conducted using the adapted redundant conditions, showing significant main effects of Condition [$F(3,90)=23.81$, $p<.001$] and of SOA [$F(2,60)=3.84$, $p<.05$], as well as a significant interaction between Condition and SOA [$F(6,180)=19.20$, $p<.001$).

SOA1=−100 ms CONDITION	predicted	obtained	t(30)	p-value	corr
Aa	394	375	−5.80	<.001	.90
Au	399	408	2.51	<.05	.93
Ua	406	410	.85	ns	.85
Uu	413	402	−2.90	<.01	.85
SOA2=0 ms CONDITION	**predicted**	**obtained**	**t(30)**	**p-value**	**corr**
Aa	358	334	−8.13	<.001	.92
Au	377	379	.22	ns	.92
Ua	364	364	−.15	ns	.85
Uu	387	369	−5.16	<.001	.90
SOA3=100 ms CONDITION	**predicted**	**obtained**	**t(30)**	**p-value**	**corr**
Aa	385	370	−3.50	<.001	.89
Au	420	403	−4.74	<.001	.93
Ua	386	409	4.45	<.001	.87
Uu	424	399	−7.39	<.001	.92

Table 4.6. Mean RTs (in ms) for the four redundant conditions Aa, Au, Ua and Uu at three SOAs as predicted by an independent separate activation model and as obtained, and t-tests and correlations for the predicted and obtained means. All correlations significant at p<.001.

For the adapted data, I next wanted to test the various congruent and incongruent redundant conditions against each other for each SOA. Since the number of such comparisons in this experiment (6 per SOA) was much higher than in Experiments 4 and 5 (3 per SOA), I decided to perform Newman-Keuls analyses with an alpha of .05, instead of planned paired comparisons. For SOA1 and SOA2, the following comparisons on the adapted conditions were significant: Aa' vs. Ua'; Aa' vs. Au'; Uu' vs.

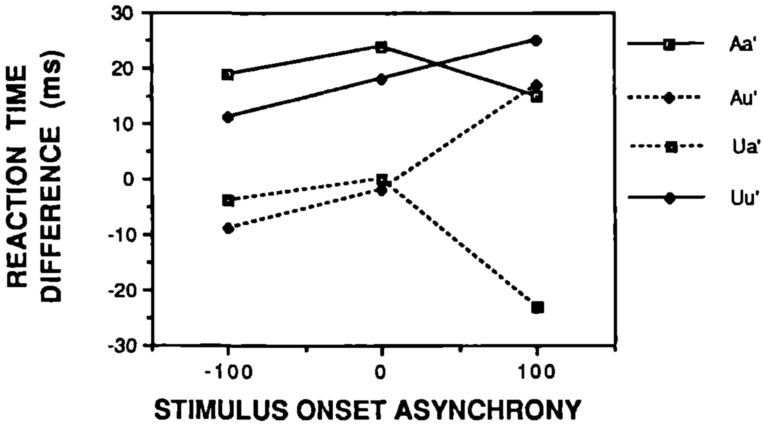


Figure 4.5. Differences (in ms) between the predicted and obtained RTs for the adapted redundant conditions over three SOAs.

Au', and Uu' vs. Ua'. For SOA3, the comparisons Aa' vs. Ua', and Uu' vs. Ua' were again significant, as was Au' vs. Ua'.

After these analyses of the adapted redundant conditions, I conducted an ANOVA on the unadapted bimodal non-redundant conditions. This analysis showed significant main effects for Condition [$F(7,210)=26.74$, $p<.001$] and for SOA [$F(2,60)=5.32$, $p<.01$], and a significant interaction between Condition and SOA [$F(14,420)=9.61$, $p<.001$]. The mean RTs over SOA for all bimodal non-redundant conditions are depicted in Figures 4.6 and 4.7 (see Appendix 2 for the associated Table). For each SOA, I tested all bimodal conditions with varying visual or auditory targets against each other and against the single-channel conditions by Newman-Keuls analyses with an alpha of .05. Of these 15 comparisons for each SOA, Table 4.7 presents the comparisons between conditions with the same target.

VISUAL TARGET	SOA1=−100	SOA2=0	SOA3=100
Ai vs. An	>	>	>
Ui vs. Un	>	>	>
Ai vs. A-	–	>	>
Ui vs. U-	–	–	>
An vs. A-	<	<	<
Un vs. U-	<	<	<
AUDITORY TARGET			
Ia vs. *a	>	–	–
Iu vs. *u	>	–	–
Ia vs. -a	–	–	–
Iu vs. -u	–	–	<
*a vs. -a	–	–	–
*u vs. -u	<	–	–

Table 4.7. Newman-Keuls analyses testing bimodal non-redundant conditions with visual and auditory targets against each other and against single-channel conditions for three SOAs. Alpha was set at .05; the direction of significant differences is indicated: e.g., > stands for first condition in comparison slower than second.

4.4.3 Discussion

When they were corrected for the effects of statistical facilitation, the results confirmed and extended those of earlier experiments for SOA1=−100 ms (visual leads by 100 ms) and SOA2=0 ms (simultaneous presentation). Again, the RTs in congruent redundant conditions (Aa, Uu) showed a much larger facilitation effect with respect to means predicted on the basis of independent separate activation than the RTs in the incongruent redundant conditions (Au, Ua). Again, the range over which there were violations

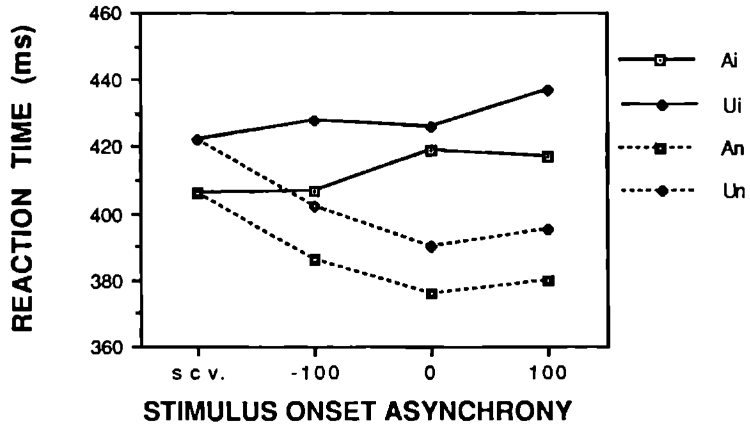


Figure 4.6. Mean RTs (in ms) to visual targets in the non-redundant bimodal conditions under three SOAs, as well as in the visual single-channel conditions. S.c.v. stands for "single-channel visual", n stands for "NOISE".

of the independent activation assumption was much larger for the congruent conditions than for the incongruent conditions. Thus, since the results with mixed stimulus presentation replicate those with blocked presentation, RT-differences between the single-channel conditions of Experiment 5 cannot have contributed to the observed differences between congruent and incongruent redundant conditions. Instead, the similarity of results provides strong support for the hypothesis of automatic cross-modal effects at a representational level, which should occur regardless of changes in the experimental design.

That the correction for the characteristics of the single-channel distributions was important, was again clearly shown by a comparison of the obtained and predicted RT-patterns for the four redundant conditions (Table 4.6): Because the incongruent Au- and Ua-conditions both had one fast channel contributing to the RT (as reflected by the single channel RTs), the raw RT-advantage of the congruent Uu-condition, built up from two slow channels, was minimal. For example, at SOA1$=-100$ ms the RT for the Ua

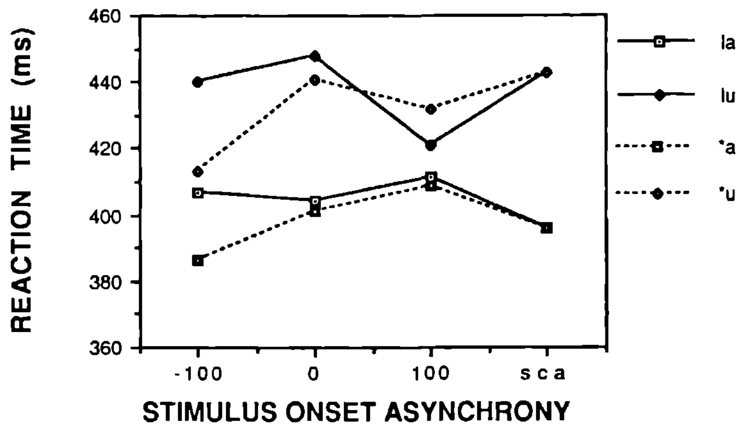


Figure 4.7. Mean RTs (in ms) to auditory targets in the non-redundant bimodal conditions under three SOAs, as well as in the auditory single-channel conditions. S.c.a. stands for "single-channel auditory".

condition was 410 ms, that for the Uu condition was 402 ms; however, the RT to the single channel /a:/ was 396 ms, but to /u:/ 443 ms! Therefore, the amount of statistical facilitation to be expected was much larger in the Ua-condition than in the Uu-condition. Table 4.6 shows this was indeed true.

The RT-pattern under SOA3=100 ms (auditory leads by 100 ms) seemed to differ from the other patterns obtained so far. Since the auditory stimulus preceded the visual here, it seemed reasonable to compare RTs for Aa with those for Ua, and RTs for Uu with those for Au, since these pairs of congruent and incongruent redundant conditions share their auditory stimulus. After correction for the characteristics of the contributing single channels, a larger facilitation effect was observed for the congruent Aa'-condition than for the incongruent Ua'-condition. Though the congruent Uu'-condition showed a slightly larger facilitation effect than the incongruent Au'-condition, this difference did not reach statistical significance (from Table 4.6: Aa' 15 ms vs. Ua' -23 ms; Uu' 25 ms vs. Au' 17 ms).

The results for the Aa-condition suggest that the auditory phoneme /a:/ was cross-modally activated by the corresponding grapheme A. The existence of grapheme-to-phoneme activation is also supported by the experimental results presented in Chapter 3 and by the experimental literature (e.g., Van Orden, 1987). (Notice that, though the race model could not be rejected for Aa under SOA3, this does not preclude the presence of coactivation). The existence of such facilitation effects, both under SOA1 and SOA3, seems to imply bidirectional cross-modal activation spreading, or use of a sublexical code common to visual and auditory processing.

It is hard to explain why RTs in the incongruent redundant conditions deviated more from the predicted RTs for SOA3 than for SOA1 and SOA2. One explanation is in terms of the division of a subject's reactions over the visual or auditory component targets. Suppose the subject had a general tendency to react to the visual signal more often than would be expected on the basis of an independent race. Such a tendency has been referred to in the literature by the term "visual dominance" (e.g., Posner, Nissen, & Klein, 1976). This would lead to longer RTs than expected on the basis of statistical facilitation when the auditory stimulus precedes the visual one, and the more so when the visual stimulus arrives later (remember that the RT is measured from the onset of the first target stimulus). Thus, the Ua-condition should suffer more than the Aa-condition, since the single-channel RTs indicate that the visual stimulus U is processed slower than A. This is in agreement with the data. However, since visual dominance leads to interference effects, one would expect RTs always to be slower than those expected on the basis of a race: Therefore, the large relative facilitation of the Au-condition is not accounted for. I will come back to this issue in Chapter 5.

A different explanation is that deviations from the independence prediction arose because of non-specific coactivation effects that were dependent on the temporal overlap of two processes (e.g., due to arousal). Taking into account the SOA of 100 ms, it was indeed found that the overlap of the visual and auditory component distributions was much larger in the Au- and Uu-conditions than in the Aa- and Ua-conditions. This would explain the larger facilitation effect in the first mentioned conditions. However, since this type of explanation basically is facilitatory in nature, it has trouble explaining the large relative inhibition in the Ua-condition. Therefore, neither of the two explanations offered here seems sufficient to account for the whole pattern of results.

Compared to the mean RTs predicted on the basis of independent separate activation, only small inhibition effects that did not seem to depend on SOA showed up in the redundant conditions. This provides some evidence for the absence of cross-modal representational inhibition. This conclusion would have been stronger, however, had the experiment included a redundant condition with a no-letter target (like the * in Experiment 5), since now it cannot be completely excluded that facilitatory coactivation effects at decision or response levels hide inhibitory effects at a representational level.

When the RTs in the redundant conditions (Figure 4.4) are compared to those in the non-redundant conditions (Figures 4.6 and 4.7), it is immediately clear that the non-redundant conditions were much less sensitive to the SOA-manipulation than the redundant conditions. While the redundant conditions consisted of two target stimuli in different modalities, the non-redundant consisted of a target stimulus in one modality and a neutral stimulus in the other. The insensitivity of the non-redundant conditions to SOA was therefore to be expected, since the neutral stimulus should have only indirect effects on the responding (e.g., via arousal – motor coactivation): No effects of statistical facilitation can occur when only one target is involved.

As can be seen in Figure 4.6, the RT-differences between non-redundant conditions with auditory /i/ and those with NOISE were larger when these neutral stimuli arrived earlier with respect to the visual target. While NOISE had general facilitatory effects with respect to the single-channel conditions, the effect of /i/ was slightly inhibitory. Figure 4.7 indicates there were RT-differences between non-redundant conditions with visual I and * as well, but only when these neutral stimuli preceded the auditory targets.

Over-all, the effects of a neutral stimulus on the target seemed to be stronger and more stimulus dependent when the neutral stimulus was auditory than when it was visual. The following account is in line with the general pattern of results. Auditory stimuli characteristically cause relatively larger arousal effects than visual stimuli (Keuss, 1987). Therefore, larger facilitation effects with respect to the single channel are expected for the NOISE-conditions than for the *-conditions. However, in order to reject a stimulus as a possible target, a subject must at least perform a partial analysis of it. It seems plausible that the amount of analysis required depends on the similarity of the stimulus to a target. Since NOISE is very different from speech, the NOISE-signal can quickly be rejected as a possible target. Relative to NOISE, a star stimulus will take longer to reject, since there is nothing *inherent* in a star that distinguishes it from a target letter. If the stimulus is the letter or speech sound "I", it may be still harder to reject as a target. According to this view, interference occurs in the "I"-conditions because the neutral stimulus must be processed to a certain extent before the attention can be shifted to the other channel. This interference effect hides the general facilitation caused by an auditory stimulus in case of the neutral /i/.

4.5 General discussion

In this chapter I investigated three issues concerning the relationship between grapheme and phoneme representations. First, can cross-modal influences between such representations be demonstrated? Second, can such cross-modal influences be only facilitatory or also inhibitory? Third, are cross-modal influences from the visual to the auditory modality and vice versa similar in size and time-course?

These issues were explored by means of the bimodal detection task, well-suited for this purpose since it yields fast RTs and allows for, but does not require, on-line cross-modal influences. In three go/no-go vowel-detection experiments, subjects reacted to specific letters (or symbols) and/or speech sounds in visual, auditory or bimodal stimuli. In some bimodal conditions, both a visual and an auditory target appeared, that were either congruent (e.g., visual A and Dutch auditory /a:/), or incongruent (e.g., visual A and auditory /u:/). To evaluate the temporal nature and the directionality of cross-modal representational contacts, the onset asynchrony of the component stimuli was varied.

In order to account for processing differences between the various visual and auditory target stimuli involved in the redundant (two-target) conditions, the RTs of these conditions were not compared directly, but only after a correction that took into account statistical facilitation expected on the basis of an independent race between the visual and auditory channels (cf. Appendix 1). The experimental results for the redundant conditions indicated that statistical facilitation indeed played an important role: Under some circumstances statistical facilitation effects were found of up to 30 ms. The importance of statistical facilitation was also apparent in the high correlations (around .90) between the means predicted on the basis of the race model and those obtained.

After the correction method was applied, RT-data from three experiments showed facilitation effects for congruent redundant conditions compared to incongruent conditions, both when the visual and auditory stimulus were presented simultaneously and when the visual stimulus preceded the auditory one by 100 ms. The presence of facilitation effects in RTs of about 330 ms under this last mentioned SOA is evidence in favor of early cross-modal facilitation at a representation level, especially when the time needed to program and execute the response is taken into account. The speed of the RTs further suggests that the effects were not under strategic control of the subject, i.e. that they were automatic (Posner & Snyder, 1975a, b). This conclusion is further strengthened by the fact that consistent facilitation effects were found in experiments with both mixed and blocked designs, and with different instructions and stimulus conditions.

Considering the second issue raised above (inhibition across modalities), not much evidence was obtained in the redundant conditions in favor of cross-modal inhibition, nor of mediated inhibition (e.g., grapheme A inhibiting phoneme /u:/ via activation of phoneme /a:/). The &- or *-conditions and the incongruent redundant conditions showed very similar RT-patterns. It could be argued that these symbols were in most respects similar to letters and thus themselves capable of cross-modal inhibitory influences. If this were so, however, one might expect a dependence of the size of the inhibitory effect on SOA (i.e., more inhibition for SOA2=0 ms, where distributions overlap more), which was not obtained.

The longer RTs found in Experiment 6 for the bimodal non-redundant conditions containing a visual target and an auditory non-target /i/ compared to the single-channel and NOISE-conditions were not considered strong evidence for inhibition effects either. It seemed more likely that this RT-difference reflected the extra processing time needed to reject a linguistic stimulus as a possible target compared to a NOISE-stimulus. The (smaller) interference effects in the corresponding bimodal non-redundant conditions with a visual neutral stimulus may be explained in a similar vein.

Finally, it is possible to ascribe the lack of inhibition in the redundant conditions to the nature of the detection task I used: If significant cross-modal or mediated inhibition occurs only after a stimulus has been identified, then a go/no-go task which requires a reaction to the first target stimulus to be detected will show little or no inhibition of either kind.

The third issue addressed in this chapter was whether cross-modal influences from the visual to the auditory modality are similar in size and time-course to those in the other direction. A comparison of the bimodal redundant RTs obtained under the two extreme SOAs (SOA1=−100 ms and SOA3=100 ms) is relevant here, as is a comparison of non-redundant conditions that have the target in different modalities.

In the bimodal redundant conditions, fast cross-modal activation effects were found both from the auditory to the visual modality (SOA1, letter leading by 100 ms), and from the visual to the auditory domain (SOA3, letter following by 100 ms). For SOA3, significant effects were found only between the grapheme A and the vowel /a:/, but the experiments reported in Chapter 3 also confirm the existence of grapheme-to-phoneme activation. However, the RT-pattern at SOA3 differed from those at SOA1 and SOA2 (simultaneous presentation). This suggests some kind of processing difference between conditions where an auditory target precedes a visual one and where it follows it. It may be remarked here that in a sense SOA1 and SOA3 are not directly comparable because of the characteristics of visual and auditory stimuli. The visual stimulus appears at once and all information needed to identify it is present from the same moment in time onwards; the auditory stimulus, however, builds up over time, and when it can be identified may vary with its temporal characteristics. This fact causes synchronisation problems for any kind of on-line cross-modal research, be it with words or with letters.

If visual and auditory stimuli have different general cross-modal effects, this should come out in different RT-patterns for non-redundant conditions with visual and auditory targets. Indeed, quite different RT-patterns were found for non-redundant conditions with visual and auditory targets. The results were shown to be compatible with an account that assumes, first, larger arousal effects caused by auditory than by visual stimuli (cf. Keuss, 1987), and, second, differences in the time to reject various non-targets because they are more or less similar to targets. For example, one might be faster to distinguish white noise from an auditory target vowel, than a star from a target letter. If this account is correct, the asymmetric results at SOA1 and SOA3 in my experiments were not due to an asymmetry on representation levels, but to differences in the characteristics of visual and auditory stimuli, or of their combination (e.g., due to visual dominance, cf. Posner, Nissen, & Klein, 1976).

The three fundamental issues concerning the relationship between grapheme and phoneme representations that have just been discussed, have consequences for all models that use such representations as mediators in higher level processes (e.g., word recognition). I will postpone a more general theoretical discussion of these consequences till Chapters 6 and 7; here I will evaluate the implications of the experimental results for the time-course model of visual word recognition, exposed in the Introduction of this chapter. This model seems to lead to several predictions that agree with the data. Consistent with the model are the findings of cross-modal facilitation effects, but no clear inhibition effects. If the argument concerning the bidirectionality of the observed effects (from visual to auditory and vice versa) is accepted, the model would also be consistent with it (though the implemented version of the model should still be modified to include phoneme-to-grapheme activation). However, unexplained by the model is the asymmetry in results depending on the modality in which the first stimulus is presented. As argued before, it seems probable that the observed asymmetry effects are related to general characteristics of visual and auditory stimuli and to the specific interaction of the two modalities. Being a model of visual word recognition, the time-course model of course leaves the issue of bimodal presentation effects untouched. Finally, and related to this point, the model does not incorporate any mechanism that explains the different influence of visual and auditory neutral stimuli in the non-redundant conditions.

To conclude, the evidence presented in this chapter supports the theoretically relevant conclusion that cross-modal facilitation between grapheme and phoneme representations occurs rapidly and automatically, and probably in both directions. The existence of such sublexical cross-modal influences must be taken into consideration by models of word recognition in both the visual and the auditory domain.

5

Bidirectional grapheme-phoneme activation in a modality decision task

How the human being processes multimodal signals has been studied from two main perspectives. Researchers interested in the attentional aspects of multimodal perception have investigated how performance varies with the division of attention among several modalities (Martin, 1980; Stanislaw, 1988). Others, especially psycholinguists, with interest in the cross-modal effects of multimodal processing have examined how specific (linguistic) stimulation in one modality affects perception in the other.

The research concerning grapheme context effects on phonemic processing (Chapter 3) and mutual grapheme-phoneme activation (Chapter 4) can of course be seen as representing this latter orientation. In Chapter 4 I tried to obtain evidence for grapheme-to-phoneme and phoneme-to-grapheme activation in a bimodal vowel-detection task by manipulation of the temporal relationship between visual and auditory target presentations. I assumed that a subject reacts more often to the first presented target in a bimodal redundant (two-target) combination. Since in a congruent redundant condition the second arriving target should cross-modally activate the first at a representation level, the RTs to such first presented targets should be facilitated relative to those in the incongruent conditions, where no such cross-modal activation takes place. Indeed, both when the visual signal preceded the auditory and when it followed it by 100 ms, the obtained RT-patterns seemed to reflect cross-modal activation effects. This suggests that cross-modal activation is bidirectional in nature.

However, the validity of the assumption that subjects reacted more often to the first presented target was questioned in Experiment 6 (Chapter 4) by a puzzling deviation of the RT-patterns from those predicted by a race model when the auditory stimulus preceded the visual by 100 ms (SOA3). If manipulation of SOA did not lead to differences in the proportion of reactions to one modality or the other, the conclusion that cross-modal representation effects are bidirectional is no longer certain.

In this chapter an experiment is presented, that was intended, first, to investigate a possible cause for the deviating RT-pattern. Second, it was meant to test the hypothesis of bidirectional activation by looking for cross-modal effects in separate responses to visual and auditory targets, and not only via the manipulation of SOA. After discussing a possible origin for the deviating RT-pattern, I will describe the bimodal task used for this purpose.

5.1 Visual dominance

One (partial) cause for the deviating RT-pattern when the auditory stimulus preceded the visual could be a tendency of subjects to react more often to the visual target stimulus than would be expected on the basis of a race (cf. the general discussion of Chapter 4). Indeed, the literature on attentional aspects of multimodal perception refers to a tendency to react predominantly to the visual modality in bimodal situations, a phenomenon that has been called "visual dominance" or "visual capture" (Posner, Nissen, & Klein, 1976).

The phenomenon of visual dominance was discovered during the second half of the last century, at the dawn of experimental psychology, by astronomers measuring the movements of stars around the heavens. When they judged the number of clock beats occurring while a star traversed a given spatial distance, their judgments disagreed very often. It turned out, that when a person was attending to a visual channel, his response to a simultaneous sound (e.g., a tone) often appeared to be delayed. The temporal order judgment of the subject was seen to be biased in favor of the light. This observation was included by Titchener (1908, p. 251) as the "Law of Prior Entry" (defined by Wundt) among his seven Laws of Attention.

Visual dominance has been demonstrated to be quite pervasive in various bimodal experiments, but some striking examples are given by Colavita (1974). Colavita had subjects match an auditory stimulus (4000 Hz, 65 dB) and a visual stimulus (6W) for subjective magnitude. Subsequently, these stimuli were presented in a mixed forced two-choice experiment (tone or light key). In some trials tone and light were "accidentally" presented at the same time. RTs to such conflict trials and single-channel visual and auditory trials were equally short (about 300 ms), but in conflict trials subjects responded almost exclusively to the light (49 out of 50 cases). Often (in 16 out of 49 cases) subjects even seemed unaware that there had been a simultaneous presentation of the tone, though light and tone were matched in subjective intensity, and the tone alone led to shorter RTs than the light in a simple single-stimulus RT-paradigm (179 vs. 197 ms). In other experiments Colavita again found a consistent prepotency of the visual stimulus over the auditory stimulus in conflict trials, even when the subject was informed about their existence, when he was specifically instructed to respond to the tone on such trials, or when the tone was subjectively twice as intense as the light. (However, later research by Egeth & Sagar, 1977, has shown several restrictions to the occurrence of visual dominance).

A theory of visual dominance was developed by Posner, Nissen, and Klein (1976). According to this theory, visual stimuli are not as automatically alerting as stimuli in other modalities (cf. Keuss, 1987), and a visual event can serve as an effective alerting stimulus only if it is first processed by "active attention". Active attention given to a modality leads to a reduction of the attentive mechanisms to input from other modalities. To compensate for the low alerting capability of visual signals, subjects would exhibit a general attentional bias towards the visual modality whenever they are likely to receive reliable input from that modality. Since the visual stimuli presented in the experiments of Chapter 4 were unambiguous and task relevant, the occurrence of visual dominance in my experiments would seem to follow from Posner et al.'s theory.

Indeed, one of the experiments in which Posner et al. found support for their theory produced results that resemble those obtained in Experiment 6 (Chapter 4) for the bimodal non-redundant conditions. Their experiment investigated the effects of auditory and visual non-target stimuli on the detection of a visual letter target ("X") or an auditory tone target. A reduction of the RT by 40 ms was obtained when white noise was added to the letter stimulus, and a reduction of 12 ms when a square flash of light accompanied the tone target. Furthermore, the auditory accessory produced a clear effect on visual processing over a range of SOAs (between -100 and 100 ms), while the visual accessory affected processing only when it preceded the target signal by 100 ms. Very similar results were obtained by me in Experiment 6 (Chapter 4). Here it was found that, when an auditory accessory (white noise) and a visual target letter occurred simultaneously, RT was reduced by about 30 ms. However, when a visual accessory (star) occurred at the same time as an auditory target, no reduction effects were obtained. The temporal range of effects for the visual and auditory accessories was similar to that of Posner et al. as well. Finally, in my experiment neutral linguistic stimuli (such as the vowel /i/) caused more interference than nonlinguistic stimuli (such as white noise). This is in line with a suggestion offered by Posner et al., that the time need to switch to the target modality depends on the depth to which a neutral stimulus is processed. To conclude this point, the task situation of the experiments in Chapter 4 was such as Posner et al. presume for the occurrence of visual dominance, and some of the RT-patterns obtained there were similar to those reported by Posner et al. for an experiment that displayed visual dominance.

To test if visual dominance in fact played a role in my experiments, I devised a speeded modality decision task, similar to that used by Colavita (1974). I reasoned that I should be able to detect visual dominance when it was known to which target modality a subject reacted to in a bimodal trial. Visual dominance would be indicated by a relatively large proportion of reactions to the visual modality at various SOAs. I therefore had subjects detect target stimuli in the visual or auditory modality as in earlier experiments, but instead of giving a simple go-reaction, they had to determine in which modality (visual or auditory) the target was presented. Thus, apart from identifying a target, the subject also had to determine the modality in which it was presented and relate the modality label to the appropriate response button. Redundant conditions were included in which two targets appeared (e.g., Aa or Au).

5.2 Bidirectionality of cross-modal representation effects

The modality decision task seems well-suited not only to test for visual dominance, but also offers an interesting alternative to SOA-manipulation for a test on the bidirectionality of cross-modal representation effects. Indeed, phoneme-to-grapheme activation should facilitate the subject's RT to a visual target when it is accompanied by a (not responded to) congruent auditory target, while grapheme-to-phoneme activation should facilitate the RT if the congruent auditory target was reacted to. Evidence for facilitation should follow, as before, from a comparison between the RTs in congruent and incongruent redundant conditions.

Since congruent and incongruent redundant conditions consist of different combinations of stimuli, the RTs to such conditions should again be corrected for differences in statistical facilitation (cf. Chapter 4). For this modality decision task, it is possible to predict the RTs expected on the basis of a race model for visual and auditory modality reactions *separately*. In a bimodal trial, the probability that at time t a reaction is given to a certain modality here is conditional upon processing in that modality having finished before that in the other modality. For example, for a reaction to the visual stimulus in a bimodal trial:

$$P(RT_{v-in-va}=t) = P(RT_v=t|RT_v<RT_a)$$

For the discrete situation, the bimodal probability in the definition can be related to the single-channel proportions by the following equation (in which RT_v and RT_a stand for reactions to the visual and auditory channel, and t stands for time):

$$(1) \quad P(RT_v=t|RT_v<RT_a) = \frac{P(RT_a>t)P(RT_v=t)}{\sum_{t'=0}^{\infty} P(RT_a>t')P(RT_v=t')}$$

The mean RT to the visual modality is equal to

$$\overline{RT}_{v-in-va} = \sum_t tP(RT_v=t|RT_v<RT_a).$$

The proportion of reactions to the visual modality is equal to

$$P(RT_v<RT_a) = \sum_t P(RT_a>t)P(RT_v=t).$$

The derivation of these equations is given in Appendix 1; the equations for $RT_{a-in-va}$ are analogous. Assuming that an independent race takes place, both the predicted mean RTs and proportions of reactions to each modality can be computed using the empirical single-channel distributions. For example, $P(RT_a>t)$ can be estimated by the complement of the empirical CDF for the auditory single-channel condition.

Since the proportion of reactions to each modality can be computed, it is possible to examine whether more or fewer reactions are given to the visual modality than expected on the basis of a race. If visual dominance exists, more visual reactions should be given than expected for all SOAs. Notice that this comparison of the obtained proportion of reactions to a modality and the proportion predicted provides a better test for visual dominance than an evaluation of the raw data, since it takes into account possible RT-advantages for visual stimuli.

Having discussed the general characteristics of the modality decision task, I now describe some more specific design features. In this experiment I expanded the number of SOAs from three to seven to obtain more information about the temporal aspects of the visual to auditory effects, and to see whether I could replicate the effect for SOA3 (auditory leads by 100 ms) in Experiment 6 (Chapter 4) for a larger set of SOAs in this new task situation.

To determine whether cross-modal representational effects could still be observed after large numbers of trials when the reactions of the subjects have stabilized (as they should if they occur automatically), I decided to run the experiment in a "psychophysical" fashion: obtaining many data points from a limited number of subjects (cf. Hell, 1987, pp. 55-56).

Finally, I again included bimodal non-redundant conditions in this experiment, for several reasons. First, these conditions with only one target stimulus and one neutral (non-target) stimulus force the subjects to really identify the presented stimuli and therefore they discourage guessing tendencies. Second, these bimodal conditions can be considered baselines that reflect the general effect of a neutral stimulus in a second channel on the reaction. As such they may provide evidence concerning the allocation of attention over the visual and auditory modality in the modality decision task. Third, since such bimodal conditions also were included in the detection experiments of last chapter, a better comparison of the two types of experiments is possible. For example, if the results for these conditions would be similar to those in the earlier experiments, this would perhaps suggest a similarity in terms of the subject's performance of the task and thus indicate the possibility of generalizing over the two experimental situations. (As we shall see in Chapter 7, this is also relevant for the development of a simulation model).

5.3 Experiment 7

5.3.1 Method

Subjects. Eight undergraduates at Nijmegen University, all native speakers of Dutch, were paid to participate in the experiment.

Design. The experiment was run in twelve sessions. In each session the following types of conditions were included: bimodal redundant (consisting of the presentation of two targets), bimodal non-redundant (one target and one non-target), single-channel target (one target), single-channel neutral (one non-target), and bimodal neutral (two non-targets). In all, 15 different types of stimulus presentation occurred, as shown in Table 5.1.

	TARGETS		NEUTRAL	
AUDITORY:	/a:/	/u:/	/i/	–
VISUAL				
TARGETS				
A	Aa	Au	Ai	A-
U	Ua	Uu	Ui	U-
NEUTRAL				
I	Ia	Iu	Ii	I-
–	-a	-u	-i	

Table 5.1. Stimulus conditions in Experiment 5, ordered by visual stimulus (column, first symbol) and auditory stimulus (row, second symbol). – stands for "no signal".

The bimodal redundant and the bimodal non-redundant conditions occurred under seven SOAs, with the visual stimulus presented 100, 40, or 20 ms before the auditory signal, with the same onset time as it, or 20, 40, or 100 ms after it. Redundant conditions were repeated 9 times under all SOAs, non-redundant conditions 6 times. Single-channel target conditions were also repeated 9 times in each session, as were single-channel neutral conditions. Bimodal neutral conditions were repeated 6 times for each SOA. This led to a total number of test stimuli in each session of: 4 (redundant conditions) * 7 (SOA) * 9 (repetitions) + 4 (bimodal neutral conditions) * 7 (SOA) * 9 (repetitions) + 4 (single-channel conditions) * 9 (repetitions) + 2 (single-channel neutral conditions) * 9 (repetitions) + 1 (bimodal neutral condition) * 7 (SOA) * 6 (repetitions) = 252 + 168 + 36 + 18 + 42 = 516 stimuli. Furthermore, 32 practice trials for each session were constructed, leading to a total number of trials in a session of 548.

The digitized recordings of the visual and auditory stimuli from Experiment 3 reported in Chapter 4 were used here again for the construction of the experimental tapes. The mean duration of the auditory stimuli was 320 ms, as was the presentation time of the visual stimuli.

Procedure. The experimental sessions were run over two weeks, with a maximum of two sessions a day. The order of sessions was varied over subjects.

At the beginning of the experiment, subjects read a written instruction. It indicated that in each trial either a visual or an auditory signal would be presented, or a combination of both. Dutch subjects were instructed to react to target signals, which were defined to be the letter or sound "A" and the letter or sound "U". In some trials, both a visual and an auditory target were presented. In such trials, subjects subjects had to push a response button labeled "ZIEN" ("SEE") whenever they first identified the visual target, but another button labeled "HOREN" ("HEAR"), whenever they first identified the auditory target. In other conditions, only one target appeared, either together with a non-target stimulus or alone. Here the subjects had to react as soon as possible by pushing the response button that specified the target's modality ("ZIEN" for visual, "HOREN" for auditory). Finally, subjects were told not to react to trials in which target stimuli were absent (catch trials with one or two exemplars of "I").

Four subjects reacted during all 12 sessions to the visual modality with their right index finger, and to the auditory modality with their left index finger. The other four subjects were instructed to do just the opposite. One left-handed subject was allocated to each group. Thus, handedness and response button were counterbalanced.

Each trial started with a 1000 Hz. warning signal of 200 ms duration. In the single-channel conditions (presentation of a visual or auditory stimulus in isolation), this warning signal was followed after 600 ms by the target stimulus. In the redundant trials, the visual stimulus followed the warning signal after 500, 560, 580, 600, 620, 640 or 700 ms (depending on SOA); the auditory stimulus was always presented after 600 ms. Each 2.5 seconds after presentation of the auditory signal a new trial was initiated.

Each session consisted of 32 practice trials followed by 3 blocks of 172 test trials. After the practice set there was a short pause in which there was an opportunity for asking questions. Each session lasted for about 50 minutes, with 3 3-minute breaks before and between blocks.

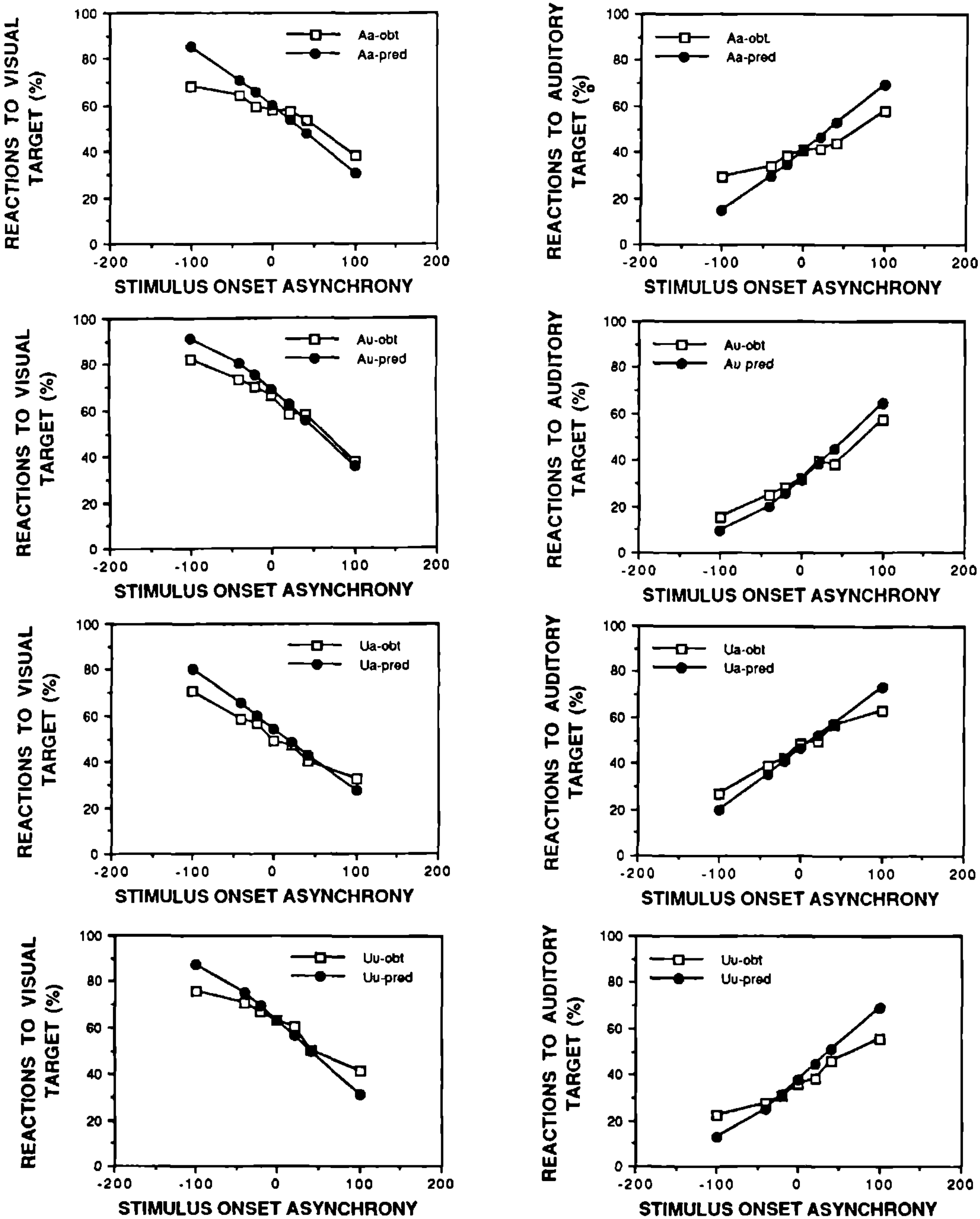
5.3.2 Results

As expected on the basis of a pilot study, large learning effects were found over the first couple of sessions. Table 5.2 illustrates the decrease in RTs and reduction in number of errors over sessions by means of the raw results for the combined bimodal non-redundant conditions.

Session	RT	Error	Session	RT	Error
1:	772	18.7	7:	433	9.2
2:	567	11.0	8:	424	8.7
3:	507	10.9	9:	427	6.4
4:	477	9.4	10:	432	6.3
5:	457	7.1	11:	419	6.4
6:	450	7.2	12:	411	5.7

Table 5.2. Mean RTs and error percentages to the combined bimodal non-redundant conditions over 12 sessions. Data outside the range 150 and 1500 ms were replaced by the mean of the appropriate condition; such outliers and reactions to the wrong modality were scored as errors.

To avoid the extra variance and possibly contaminating consequences due to learning in the starting sessions of the experiment, the first two sessions of each subject were considered practice sessions, and only the test trials of the last 10 sessions were further analysed (cf. Hell, 1987). Mean


Aa-obt
Aa-pred
REACTIONS TO VISUAL TARGET (%)
STIMULUS ONSET ASYNCHRONY
Aa-obt
Aa-pred
REACTIONS TO AUDITORY TARGET (%)
STIMULUS ONSET ASYNCHRONY
Au-obt
Au-pred
REACTIONS TO VISUAL TARGET (%)
STIMULUS ONSET ASYNCHRONY
Au-obt
Au pred
REACTIONS TO AUDITORY TARGET (%)
STIMULUS ONSET ASYNCHRONY
Ua-obt
Ua-pred
REACTIONS TO VISUAL TARGET (%)
STIMULUS ONSET ASYNCHRONY
Ua-obt
Ua-pred
REACTIONS TO AUDITORY TARGET (%)
STIMULUS ONSET ASYNCHRONY
Uu-obt
Uu-pred
REACTIONS TO VISUAL TARGET (%)
STIMULUS ONSET ASYNCHRONY
Uu-obt
Uu-pred
REACTIONS TO AUDITORY TARGET (%)
STIMULUS ONSET ASYNCHRONY

RTs and number of reactions for both modalities were now computed for each subject and each condition. Latencies greater than 750 ms or smaller than 150 ms were treated as errors. The total percentage of missing and extreme values in the eight bimodal conditions was 5.8 %. The percentage of "false alarms", i.e. reactions to the catch-trials was 3.8 % for the I-condition, 6.6 % for the i-condition, and 13.4 % for the Ii-condition.

To evaluate the presence of visual dominance, I analyzed the percentages of modality reactions. Figure 5.1 graphically represents the percentage of reactions to the four redundant conditions in each of the seven SOAs, split up in reactions to the visual and auditory modality (the sum of the obtained percentage of reactions to the visual and auditory components of a redundant condition equals the total percentage of correct reactions for that condition). Negative SOAs indicate conditions for which the visual stimulus appeared first, while under positive SOAs the auditory stimulus was leading. This Figure also indicates the percentage of reactions expected for visual and auditory targets if a race model would hold (all represented values can be found in Appendix 2). This percentage was obtained by summing over all values of t $P(RT_a>t)P(RT_v=t)$ for the visual reactions, and $P(RT_v>t)P(RT_a=t)$ for the auditory reactions (cf. Equation 1 in the introduction of this chapter and Appendix 1).

As the Figure clearly shows, for early SOAs less reactions were given to the visual modality than predicted, and for late SOAs less reactions to the auditory modality than predicted. The results for the early SOAs are clearly inconsistent with the hypothesis of visual dominance, that would predict more visual-target reactions over the whole range of SOAs. To test whether the observed differences between obtained and predicted percentages of reactions were statistically significant, and to investigate at the same time whether there were differences between congruent and incongruent conditions, the following analysis was performed.

I subtracted, for each SOA and each subject, the obtained percentages

---

Figure 5.1. Obtained and predicted mean percentage of reactions to the visual and auditory targets, and the difference between the obtained and predicted percentages, in the redundant conditions under all SOAs.

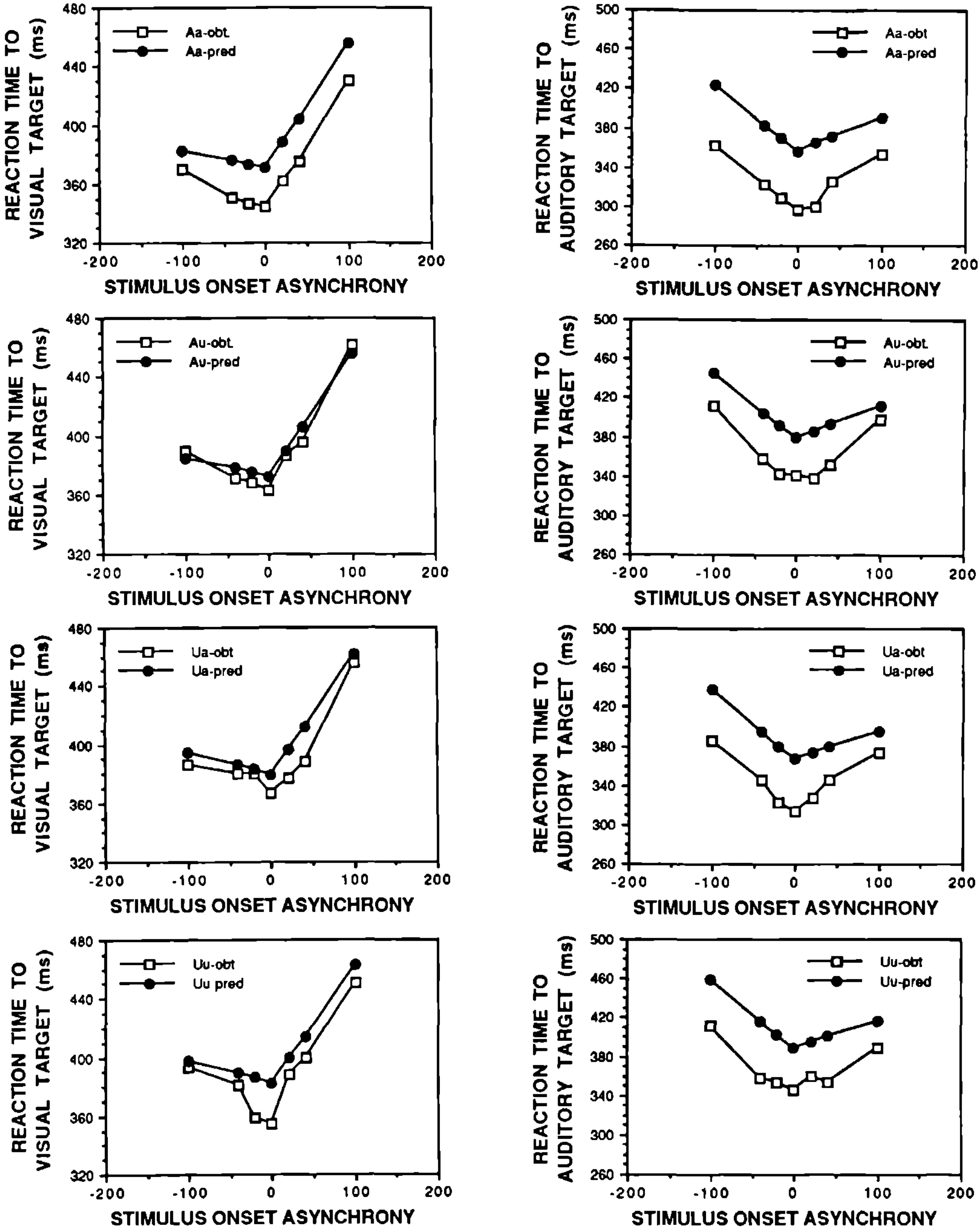
from the predicted for the visual and auditory reactions in all redundant conditions. This difference score (comparable to the differences visible in Figure 5.1) was subsequently averaged over the visual and auditory reactions for each redundant condition, and then over the congruent or the incongruent conditions. Thus, the result was, for each SOA and each subject, a difference score for the congruent conditions (Aa, Uu) and one for the incongruent conditions (Au, Ua). On these difference scores an ANOVA was performed with the factors SOA (1-7) and Congruence (congruent, incongruent). The main effect for SOA was significant [$F(6,42)=16.62$, $p<.001$], indicating that the difference scores for the obtained and predicted percentages varied over SOA. In combination with Figure 5.1 this result disconfirms the visual dominance hypothesis. Furthermore, the main effect for Congruence showed a trend towards significance [$F(1,7)=2.62$, $p=.15$], while the interaction of SOA and Congruence was significant [$F(6,42)=3.91$, $p<.01$]. Inspection of the data represented in Figure 5.1 indicates that the congruent conditions tended to show larger deviations from the predicted percentages of reactions at the extreme SOAs (especially SOA1=−100 ms, SOA5=20 ms, SOA6=40 ms and SOA7=100 ms).

I now turn to the analysis of the reaction times. Figure 5.2 shows the mean RTs to the four redundant conditions in each of the seven SOAs, split up in reactions to the visual and auditory modality. The mean RTs for the single-channel conditions (not represented in the Figure) were 397 ms for A, 413 ms for U, 434 ms for /a:/, and 456 ms for /u:/ (see Appendix 2).

To compare the RTs in the congruent and incongruent redundant conditions, the following procedure was used to take differences resulting from statistical facilitation into account. After subtraction of the appropriate SOA_v or SOA_a, depending on whether the auditory or visual stimulus was leading (cf. Method of Analysis section in Chapter 4), the RTs from the single-channel conditions were ordered in 1 ms steps for the analysis to obtain the highest resolution in the predictions. Only RTs between 150 and 750 ms were included in the analysis. The mean predicted RTs for

---

Figure 5.2. Obtained and predicted mean RTs to visual and auditory targets, and the difference between obtained and predicted RTs, in the redundant conditions under all SOAs.



REACTION TIME TO VISUAL TARGET (ms)
REACTION TIME TO AUDITORY TARGET (ms)
STIMULUS ONSET ASYNCHRONY
Aa-obt
Aa-pred
Au-obt
Au-pred
Ua-obt
Ua-pred
Uu-obt
Uu pred
Uu-pred

visual and auditory targets were computed from the visual and auditory components of the predicted minima distribution. These components were predicted using the single-channel conditions by means of the formula given in the introduction of this chapter. These mean predicted RTs are graphically represented in Figure 5.2 (see Appendix 2 for the exact values).[1]

For each subject, the mean predicted RT for a certain redundant condition under a specific SOA was subtracted from each RT obtained in that condition and with that SOA (thus, the subtraction involved different numbers of reactions, depending on the specific subject, condition and SOA). This difference score was used to test the various congruent and incongruent conditions against each other for the visual and auditory targets separately. Disregarding SOA, paired planned comparisons tested congruent redundant conditions against incongruent conditions with the same target. For the reactions to visual targets, Aa showed significantly larger facilitation effects than Au [$t(6)=11.27$, $p<.001$], but Uu did not differ significantly from Ua [$t(6)=-.02$, ns]. For the reactions to auditory targets, Aa showed significantly larger facilitation effects than Ua [$t(6)=4.29$, $p<.01$], as did Uu compared to Au [$t(6)=3.31$, $p<.05$]. Therefore, with the exception of visual reactions to U, larger facilitation effects were found for the congruent conditions, which constitutes evidence for coactivation effects at a representational level in these conditions.

Apart from these analyses on the adapted redundant conditions, we performed an ANOVA on the data of the bimodal non-redundant conditions. This analysis showed marginally significant main effects for Condition [$F(3,21)=3.2$, $p<.05$] and SOA [$F(6,42)=2.26$, $p=.06$) and a significant interaction between Condition and SOA [$F(18,126)=5.88$, $p<.001$]. Figures 5.3 and 5.4 show the mean RTs for neutral conditions with a visual or an auditory target, together with the corresponding single-channel RTs.

5.3.3 Discussion

Since I do not know of similar experiments with the modality decision task, I will first evaluate how sensitive the task was with respect to SOA and stimulus characteristics before I discuss the analyses of the data. The task turned out to be more difficult to perform than we expected. In the first few sessions, subjects reported that they often knew they had to react, but knew not which of the two response buttons indicated the right response. Since neither identification of a target, nor relating an identified target to one of two responses is problematic (cf. the auditory vowel-detection task

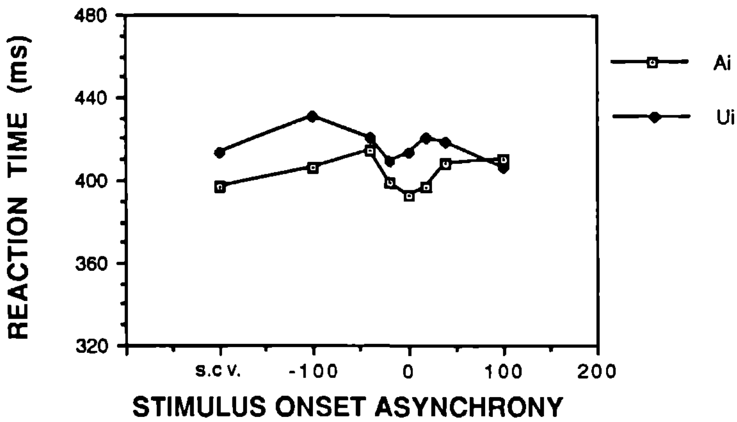


Figure 5.3. Mean RTs (in ms) in the two bimodal non-redundant conditions with visual targets under seven SOAs. Single-channel RTs to visual targets (s.c.v.) are also indicated.

used in Chapter 3), it seems that subjects at first had trouble retrieving the modality label of a target. (Of course, outside the task situation making a distinction with respect to modality is not often necessary). However, after a while the task became automatized, and RTs became short while the number of errors decreased.

To be able to test my hypotheses, the RTs should become sensitive to the characteristics of the targets presented and their combination at a certain onset asynchrony. Figures 5.1 and 5.2 showed that the subjects' reactions varied both in number and RT with the specific combination of stimuli (Aa, Au, Ua or Uu) and their temporal relationship (the number of reactions to a stimulus increased as it was presented earlier with respect to the second stimulus).

Since in redundant conditions reactions to both targets could in principle be correct, the procedure used did not allow me to distinguish intended or wrong reactions here. However, the deviation between predicted and obtained percentages of reactions to a modality can be considered as a measure of the subject's sensitivity with respect to SOA and stimulus characteris-

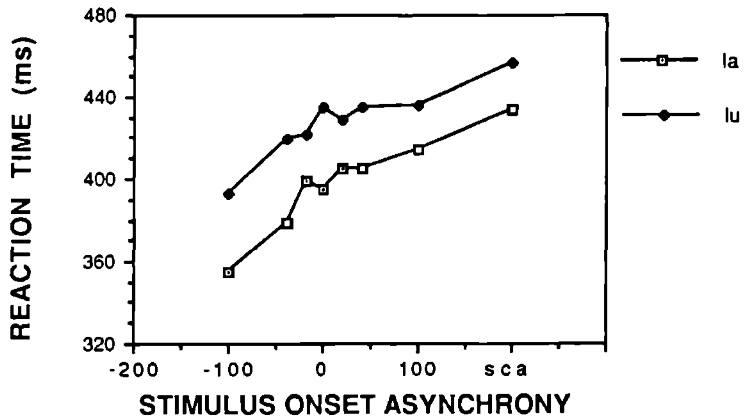


Figure 5.4. Mean RTs (in ms) in the two bimodal non-redundant conditions with auditory targets under seven SOAs. Single-channel RTs to auditory targets (s.c.a.) are also indicated.

tics. Complete guessing or insensitivity, for example, would lead to 50% of reactions to each modality and thus to large deviations from the predicted percentage. Because the deviation seldom exceeded 10%, and the over-all percentage of errors in the experiment stayed within reasonable bounds, I assume there is only limited noise in the data.[2]

Having established the sensitivity of the task to both target characteristics and SOA, I proceed to discuss the data in terms of percentage and speed of modality reactions. The presence of visual dominance in the data was evaluated by analyzing the percentage of reactions to each modality. Because the over-all percentage of reactions to the visual modality was more than 50%, one might be tempted to interpret the results as evidence for visual dominance. However, compared to the percentage of reactions predicted on the basis of the relative speed of the single-channel conditions, the results did not show any evidence in favor of visual dominance. When the visual stimulus preceded the auditory, more reactions were given to the visual modality but still less than would be expected on the basis of a race. These results therefore seem to deviate from those of Colavita (1974) who found visual dominance even when the auditory signal was subjectively made twice as intense as the visual.

If this finding with the modality decision task may be generalized to the bimodal vowel-detection task of Chapter 4, the deviating pattern of results for SOA3 (auditory leading) in Experiment 6 of that chapter (see introduction of this chapter) cannot be interpreted in terms of visual dominance. Furthermore, in contrast with that experiment, the RTs in the incongruent conditions here did not seem to behave differently over SOA. While the deviating RT-pattern was not replicated here, the differences between obtained and predicted RTs did show some fluctuations. Perhaps the deviations in Experiment 6 could be ascribed to similar, larger RT-fluctuations (the number of replications per item in Experiment 6 was much lower than in the earlier experiments, namely 20).

Apart from the percentage of modality reactions, the speed of reactions to different redundant conditions was analyzed to test the hypothesis of cross-modal contacts between grapheme and phoneme representations. As in earlier experiments, more facilitation of RT with respect to the prediction of a race model was generally found in congruent conditions than in incongruent conditions. The results for the Aa-condition further suggest that cross-modal activation was bidirectional, i.e., spread both from the visual to the auditory, and from the auditory to the visual modality. The evidence for these hypotheses was weaker for the Uu-condition. (Perhaps this could be taken as evidence that cross-modal activation strength depends on stimulus characteristics such as frequency, since the letter U is much less frequent than the letter A). In combination with the results for the three SOAs in Experiment 6 (Chapter 4), these data provide strong support the hypothesis of fast mutual cross-modal activation between graphemes and phonemes.

One result that further strengthens this conclusion is the interaction between Congruence and SOA for percentage of reactions to a modality. More reactions were given to the auditory target than predicted when the visual target was leading, but the deviation was larger for the congruent than for the incongruent conditions. Yet the same stimuli were involved in both types of conditions, only in different combinations. The following explanation of this result is in complete agreement with the hypothesis of early grapheme-to-phoneme activation. The grapheme corresponding to the leading visual target quickly activates the associated phoneme. When an auditory target now arrives, its phoneme representation will already be partially activated in case of the congruent conditions, but not in case of the incongruent conditions. Thus, the phoneme will be identified faster in the congruent conditions, which results in relatively more responses to the auditory modality. Since the extent to which the phoneme is pre-activated depends on the amount of time between visual and auditory target presentation, larger deviations should be found for larger SOAs, as is empirically confirmed. Analogously, the result for late SOAs (relatively more reactions to the visual target when the auditory is leading) can be interpreted as evidence for fast phoneme-to-grapheme activation.

The results obtained in this experiment incite to continue research by means of the modality decision task. However, before modality decision can be established as a new experimental paradigm, several questions with respect to performance of the task should be answered.

First, why were the reactions to the single-channel conditions in this experiment so slow relative to those in the bimodal conditions? One possibility is that subjects (as they sometimes suggested on the basis of their experience) waited for the appearance of a second signal in the single-channel conditions.[3] Since the method to predict the RTs in bimodal conditions uses the obtained single-channel distributions, such a tendency would lead to a prediction of too long bimodal RTs.

Second, did any response competition occur between the visual and auditory targets in the modality decision task? On the basis of Experiment 3 (Chapter 3) one would expect such response competition effects. As will be remembered, in that experiment strong inhibition effects relative to a bimodal baseline were found when a visual accessory was indirectly associated with a different response than the auditory target. A comparison of the bimodal non-redundant and redundant conditions in the modality decision task (see Figures 5.2, 5.3, and 5.4) indeed suggests that some response competition was present. At the extreme SOAs of -100 and 100 ms, the RTs were shorter in the bimodal non-redundant conditions than in the redundant conditions, even though the RTs in the redundant conditions are statistically facilitated. If response competition effects exist, they may induce larger deviations between predicted and obtained bimodal RTs.

Third, did other non-controlled and non-representational effects (such as allocation of attention) influence the results in the redundant conditions? The bimodal non-redundant conditions provide some evidence that points in that direction. The RT-curves for these conditions were rather flat over the SOA-range when the target was visual, but increased more than 40 ms when the target was auditory! One explanation to be tested is that the subjects in first instance directed their attention mainly to the visual modality (since in the course of the experiment they "learned" that the visual stimulus could appear at the shortest delay after the warning beep), and then shifted their attention to the other modality if the visual stimulus turned out to be a non-target. This type of strategy could also influence the processing of the component stimuli in the redundant conditions.[4]

Whether this explanation turns out to be correct or not, the differences between the bimodal non-redundant conditions here and in Experiment 6 (Chapter 4) suggest there were performance differences between the modality decision task and the bimodal vowel-detection task. However, the following chapter will show that, despite possible performance differences between tasks, the results of all experiments together provide a consistent and coherent view on cross-modal contacts.

6

General discussion of the empirical results and conclusions

In the Introduction of this thesis, several questions were formulated concerning contacts of grapheme and phoneme representations during bimodal sublexical processing. The experimental studies reported in the previous three chapters were designed to answer these questions by means of three different types of tasks: two-choice auditory vowel-detection (a focused attention task), go/no-go bimodal vowel-detection (a divided attention task), and two-choice modality decision (a divided attention task). In this chapter the results across these tasks are compared to construct an over-all picture of structural and temporal aspects of grapheme-phoneme contact. All of the questions posed in the Introduction are reviewed and, and as far as is possible on the basis of the results, answered. Further, potential differences among the three task types are considered, e.g., in the allocation of attention to the visual and auditory modalities. The chapter concludes with a discussion of the theoretical consequences of the research and of the directions future research may take.

6.1 Structural and temporal aspects of grapheme-phoneme contact

1. *Does any cross-modal activation occur between grapheme and phoneme representations during the processing of visual and auditory linguistic material?*

In all three empirical chapters it was investigated whether or not cross-modal grapheme-phoneme activation occurs during bimodal processing of sublexical stimuli such as vowels, consonants and syllables. To answer this question, I included in the three types of experiments congruent redundant conditions, incongruent redundant conditions and bimodal non-redundant conditions, and compared the RTs obtained for these conditions. Congruent redundant conditions involved two nominally identical visual and auditory target stimuli, such as the letter A and the speech sound /a:/ in Dutch. Incongruent redundant conditions involved two nominally different stimuli, such as letter A and speech sound /e:/. Bimodal non-redundant conditions consisted of one neutral (non-target) stimulus and one target, such as * and /u:/. The different experiments resulted in similar RT-patterns that confirmed the hypothesis of cross-modal activation between graphemes and phonemes. Larger facilitation effects were consistently obtained in congruent redundant conditions than in incongruent redundant or bimodal non-redundant conditions.

2. *Does cross-modal activation occur both from the visual to the auditory domain, and from the auditory to the visual domain?*

The finding of both grapheme-to-phoneme and phoneme-to-grapheme activation effects leads to an affirmative answer to this question. Sublexical cross-modal activation indeed seems to be *bidirectional*, i.e., it spreads both from the visual to the auditory domain and vice versa.

Grapheme-to-phoneme activation effects were found in all experiments that searched for them (Experiments 1, 2, 3, 6, and 7), both for consonants and for vowels. The size of these effects seemed to depend on the stimuli involved. In the auditory vowel-detection task, for example, the results were somewhat stronger for "A" than for "E". When the results of the bimodal vowel-detection experiments were properly adjusted, the results were stronger for "A" than for "U". The same was true for the results for "A" and "U" in the modality decision experiment.

Phoneme-to-grapheme activation effects were consistently found in all divided attention experiments for both "A" and "U" (Experiments 4, 5, 6, and 7). In the vowel-detection experiments (Chapter 4), adjusted RTs to a visual target were faster when followed by a congruent auditory target than by an incongruent auditory target (SOA=−100 ms). In the modality decision task (Chapter 5), reactions to the visual modality were faster when the visual targets were accompanied by congruent auditory targets than when accompanied by incongruent auditory targets.

Cross-modal facilitation effects of similar *size* were found in the auditory and bimodal vowel-detection tasks. The facilitation effect for simultaneous presentation of the letter A and speech sound /a:/, for example, varied between 10 and 25 ms. In the auditory vowel-detection Experiment 3 the average RT-difference between Aa- and Pa-conditions was 17 ms, and in bimodal vowel-detection Experiments 5 and 6 the RT-difference between Aa and Ua/Au after correction for statistical facilitation was 13 and 24/26 ms (all at SOA=0 ms).

However, several empirical and theoretical considerations limit the scope of this finding. First, the size of the effects seems dependent on the particular stimuli involved, possibly because of differences in connection strength (as argued above). Second, it may also depend on differences in stimulus processing time, since stimuli that take longer to be processed are potentially subjected to cross-modal activation over a longer stretch of time. Third (and a different type of argument), it varies with SOA, since the time available for cross-modal activation increases with a larger distance between accessory and target. Finally, it must to a certain extent depend on task-type for the same reason: In auditory focused attention the available time for cross-modal activation can be varied to a much larger extent, since no response is required to the visual signal.

3. What can be said about the temporal aspects of cross-modal activation?

The results indicate that cross-modal activation spreading starts already early in processing (relative to the over-all RT). In Chapter 4 it was suggested that phoneme-to-grapheme activation effects already played a role within 140 ms of processing of the visual stimulus. A similar estimate could be given for grapheme-to-phoneme effects on the basis of the bimodal vowel-detection experiments. However, the absence of cross-modal activation effects in auditory vowel-detection Experiment 3 when the auditory stimulus was presented 100 ms before the visual signal indicates one must be cautious here.

4. Can cross-modal activation between graphemes and phonemes be only facilitatory, or inhibitory as well?

The experiments reported in this thesis did not provide convincing evidence in favor of the existence of cross-modal inhibition effects at a representational level. For the bimodal vowel-detection task, incongruent redundant conditions with letters led to similar results as those with symbols (such as * or &), but, even more important, all these conditions led to RTs that were well-predictable by a simple race model that did not assume any cross-modal inhibition. For the auditory vowel-detection task, the observed inhibitory effects for the incongruent conditions were ascribed to response competition: A grapheme activates its associated phoneme, which in turn may activate an inappropriate response unit. An argument in favor of this suggestion was that bimodal non-redundant conditions generally showed RTs that were faster than in the single-channel conditions, not slower as would be expected when some sort of cross-modal (or even intra-modal) inhibition operates.

5. *Does cross-modal activation take place automatically, i.e. fast and without conscious control of the subject?*

The answer to this question can be brief. Cross-modal activation effects occurred in all experiments, despite considerable changes in design and task requirements. In the auditory vowel-detection task, cross-modal effects occurred even though the identity of the visual stimulus was irrelevant for the decision on the auditory target. Also, reaction times in all experiments were quite fast. These observations support the hypothesis that cross-modal activation occurs automatically.

6.2 Allocation of attention in the three types of experiments

The bimodal non-redundant conditions that were included in most of the experiments were not only used in comparison with the redundant conditions, but fulfilled two other purposes. First, they were compared with the single-channel conditions to investigate the general influence of an added neutral visual or auditory stimulus on the reaction to a target signal. Second, a comparison of these conditions themselves over experiments was meant to provide information about differences between experiments in the way attention was allocated to the visual and auditory modalities. In the next section, I consider both of these issues simultaneously, as they turn out to be closely related.

Both in Experiment 3 (auditory vowel-detection) and in Experiment 6 (bimodal vowel-detection), a bimodal non-redundant condition was included in which a neutral star-stimulus preceded an auditory target by 100 ms, appeared simultaneously with it, or followed it by 100 ms. In both experiments, the RTs slowly increased over SOA (e.g., for the *a-condition: Exp. 3: 399-403-408, Exp. 6: 386-401-409). In Experiment 3, the increase in RTs over SOA in the bimodal non-redundant conditions was interpreted as the result of a non-specific influence of a visual stimulus on the processing in the auditory modality. The increase over SOA in Experiment 6 can be interpreted in the same fashion; this raises the question whether the RTs in the redundant conditions in Experiment 6 were influenced by this non-specific influence as well (e.g., in Chapter 4 Table 4.6 indicates that the obtained RTs for the incongruent conditions were often slightly slower than those predicted by a race model).

Both Experiment 3 and Experiment 6 furthermore included bimodal non-redundant conditions with visual neutral letter stimuli (letters P and I) and auditory target vowels. Again, similar curves were obtained for both experiments, but with a less pronounced increase in RTs over SOAs (from -100 to 100) than for the star-conditions (cf. Chapter 3, Figure 3.3, and Chapter 4, Figure 4.7). Because of the variation in the RTs of different non-redundant conditions, it is hard to decide whether this difference between star- and letter-conditions has any theoretical significance or not. Experiment 7 (modality decision) also included a bimodal non-redundant condition with a letter (I). For this condition, the experiment differed markedly from the experiments just mentioned: Its results showed a steep increase over SOA (cf. Chapter 5, Figure 5.4). However, for the other bimodal non-redundant condition included in Experiment 7 (visual target with auditory neutral /i/) the results were more similar to the comparable condition in Experiment 6 (cf. Figures 5.3 and 4.6).

These results as a whole suggest that the influence of a non-target stimulus on the RT to the target stimulus depends to a certain extent on the specific task situation. One obvious difference between the tasks seems to lie in the allocation of attention to the visual and auditory modalities. A priori, one would expect that in the focused attention experiments (auditory vowel-detection) most of the attentional capacity was assigned to the auditory channel; in the divided attention experiments (bimodal vowel-detection and modality decision), it was probably allocated more evenly to the visual and the auditory channel. This leads to the expectation that a neutral visual stimulus will interfere more with auditory processing in the divided attention task than in the focused attention task (relative to a single channel), since it can claim relatively more attention in the first mentioned task (cf. Johnston & Wilson, 1980, for a similar account based on a comparison of focused and divided attention results). The results for the bimodal non-redundant conditions in the two task types corroborate this view. When the RTs in the star-conditions are subtracted from those in the single-channel conditions for the same SOAs (combining the two auditory targets), a positive RT-difference results in auditory vowel-detection Experiment 3, but a small negative difference in bimodal vowel-detection Experiment 6. The same is true for, respectively, the P- and I-conditions (combining again the two auditory targets).

Since the main interest in my experiments concerned the structural and temporal aspects of cross-modal activation, I did not investigate the attention issue further. Most relevant, however, is the assumption that in the experiments there were no differences between the congruent and incongruent conditions in terms of attention allocation. For auditory vowel-detection, it seems unlikely that presenting one visual stimulus or another (e.g., P or A) would in itself lead to differences in allocation of attention. For bimodal vowel-detection, RT-advantages were consistently found for the (adapted) congruent conditions compared to the incongruent conditions in a variety of experimental situations (with varying instructions and designs) and for several SOAs. This is evidence that the observed RT-advantages are not attributable to attentional differences between those conditions.

6.3 Theoretical consequences of the research and possible extensions

Taken together, the answers to the questions posed above support a view in which the visual and auditory sublexical processing systems are tightly connected or interrelated. Though the visual and auditory input signals vary in many respects (see Introduction), both quickly seem to activate the same orthographic and phonological representations. It is therefore perhaps more correct to conceive of the visual and auditory sublexical processing systems as subsystems of one larger linguistic system, than as autonomous systems in their own right. The close connection between visual and auditory processing systems seems to have a clear origin and a definite function. When we learn to read and write, cross-modal links between the visual and auditory processing systems are established right from the beginning: We learn and constantly reinforce certain grapheme-phoneme correspondences when we articulate what is written, or write what we hear. When we store linguistic information, we may store it in both types of format, in order to increase the chance of retrieving it later. And when we need to perform a particular task (e.g., letter search or letter recall), we may prefer to use one code or the other, whichever is more appropriate in the task context.

Graphemes and phonemes constitute word representations, and a close relation between these sublexical representations may have consequences for the way words are recognized. I first consider the relevance of the present research for word recognition models in general. Subsequently, I evaluate in which respects the obtained evidence is consistent or inconsistent with particular word recognition models. Finally, I describe several ways in which this research could be expanded and related more directly to word recognition.

To determine whether and to what extent orthographic or phonological information influences auditory or visual word recognition is a difficult task. To prove that, for example, sublexical and prelexical cross-modal activation of phonemes by graphemes affects visual word recognition, simultaneous testing of the following three hypotheses would be required:

1. Phonological information is activated during visual word recognition;

2. Phonological information is activated automatically, sublexically and prelexically, i.e. without involvement of the lexical (=word) level;

3. Phonologically activated information influences the visual word recognition process.

The visual word recognition research reviewed in Chapter 3 seems to have collected convincing evidence on the first and third point, but has had problems to prove the second point because of the confounding of orthographic and phonological factors in visual experiments (see Chapter 3). It is this second point that the present research has focused upon. Nonword syllables and single graphemes and phonemes were used to exclude the possibility that the observed phonological effects have occurred after word recognition. By applying bimodal tasks, it was made sure that the types of representations activated by cross-modal activation were indeed excitable by both visual and auditory processing systems. I have shown further that sublexical representations in the visual or auditory modality activate associated representations in the other modality quickly and automatically: The activation effects observed were found to operate in less than 200 ms of processing time.

Therefore, in combination with relevant experimental evidence obtained in the visual domain (Van Orden, 1987; Van Orden, Johnston, & Hale, 1988; Perfetti, Bell, & Delaney, 1988) and in other bimodal studies (Frost, Repp,

& Katz, 1988; Frost & Katz, 1989), my research offers convergent evidence for the existence of fast and automatic, not lexically mediated, cross-modal activation during word recognition.

Though conclusive empirical evidence has long been lacking, the existence and relevance of sublexical representations such as graphemes and phonemes during word recognition has been taken for granted in many well-known models. The results presented here are directly relevant to such models. Consistent with models such as the time-course model (e.g., Seidenberg, 1985b, 1987) and NETtalk (Sejnowski & Rosenberg, 1986), is my finding of *facilitatory* representational effects between nominally identical graphemes and phonemes (e.g., A-/a:/).

Interactive activation models of word recognition such as these can account for my finding of differences in the size of cross-modal effects for different graphemes and phonemes (e.g., "A" vs. "U") by assuming different weights on the grapheme-phoneme connections involved. These weights could depend on, among other factors, the frequencies of the grapheme and phoneme in question and on how often they were paired (cf. Seidenberg and McClelland, 1989).

However, inconsistent with NETtalk and a suggested extension of the interactive-activation model (Rumelhart & McClelland, 1982, p.89) are my findings concerning *inhibitory* effects. The experiments reported suggest that, first, cross-modal inhibition effects at a representational level do not occur (cf. Chapter 4). Second, inhibitory effects that do occur seem to be due to intra-modal response competition (cf. Chapter 3). If further research confirms the absence of both inter- and intra-modal inhibition effects at a representation level, this constrains the possible architecture of interactive activation models for word recognition that incorporate sublexical representations: Such models then should postulate inhibitory connections neither between nor within the representation systems for graphemes and phonemes (see Frauenfelder, Segui, and Dijkstra, in press, for other results in favor of this conclusion).

Interactive activation models assume that associations between (strings of) graphemes and (strings of) phonemes arise through learning and experience. In my research I have investigated one possible mapping between graphemes and phonemes that may thus develop (e.g., between the grapheme E and the phoneme /e:/). However, within words mappings between graphemes and phonemes are known to be many-to-many, not one-to-one. For example, in Dutch the letter E (letter name "ay" in Dutch) can be associated not only with /e:/ but also with /e/ or /ə/ (cf. the Dutch word HERLEZEN, pronounced as /herle:zən/). It would be reasonable to assume that all of these mappings would mediate cross-modal activation effects, though perhaps of different strength. Similarly, letter clusters could activate single phonemes, and vice versa (cf. ch and /x/). By means of the auditory vowel-detection task, which has been shown to be sensitive to cross-modal effects in this thesis (Chapter 3), such hypotheses could easily be tested experimentally by combining letter and sound strings.

Indeed, a mechanism of many-to-many mapping as I envision here has been incorporated in the recent word recognition model for Serbo-Croatian that was developed by Lukatela, Turvey, Feldman, Carello, & Katz (1989). The Serbo-Croatian language is written with two partially overlapping script systems, Roman and Cyrillic. Of the seven uppercase letters that these scripts share, four are ambiguous in that they refer to different phonemes. For example, H is read in Roman script as /xə/, and in Cyrillic as /nə/. In the connectionist model that Lukatela et al. develop, there is a double grapheme representation for such ambiguous letters, each of which is connected to its own phoneme representation. Also, different letters from the Roman and Cyrillic script can be related to one and the same phoneme representation (e.g., grapheme units B and V both connect to phoneme unit /və/). The same type of many-to-many mapping mechanism could be applied to the graphemes and phonemes within one language (e.g., Dutch).

As an extension of the present research to the word level, a more simultaneous test could be performed of the three hypotheses reported above. Especially the auditory vowel-detection task seems suited for experiments with words, since nonwords and words can be combined and relatively few replications are needed. For example, it could be investigated how letter clusters that are words or nonwords influence the auditory recognition process of congruent or incongruent words. One theoretical issue in which psycholinguists recently have become interested is which role orthographic and phonological "neighbors" play during recognition of target words. "Neighbors" are words that differ from targets in only few aspects, e.g., one letter or speech sound (Coltheart et al., 1977). It has been shown in both the visual (e.g., Grainger, O'Regan, Jacobs, & Segui, 1989; Andrews, 1989) and the auditory domain (e.g., Goldinger, Luce, & Pisoni, 1989) that the speed with which words are recognized does not only depend on characteristics of the words themselves (e.g., phonological make-up and frequency), but also on the size and structural aspects of the "neighborhood" they are in. Since my research has demonstrated a fast activation of both graphemic and phonemic information on the basis of visual or auditory input, the question arises whether perhaps both orthographic and phonological neighborhoods play a role during the recognition of any visual or auditory word (this question is analogous to that asked in visual word recognition research with bilinguals, namely whether neighborhoods of words in both languages are activated; Grainger & Dijkstra, in preparation). By clever manipulation of the neighborhood characteristics of visual and auditory stimuli, the auditory detection task might help to sort out the contribution of orthographic and phonological neighbors to the recognition of visually and auditorily presented words.

The present research has been restricted to bimodal processing in the sublexical domain, which is interesting in its own right. In the next chapter I explore how the account of bimodal processing that has been given here, can be incorporated in a computer-simulation model. Building such a model can uncover hidden assumptions and helps to make the presented view on bimodal processing more explicit.

7

A simulation model for bimodal processing

In this last chapter I present a blueprint for a simulation model intended to formalize the view on bimodal processing that was developed in the preceding chapters. Though the construction of a simulation model is a rather labour-intensive affair, it is quite profitable. It forces one, for example, to be much more explicit than a mere verbal model description requires. It helps to uncover hidden assumptions and fill in theoretical gaps. It has heuristic value in that it stimulates the generation of hypotheses regarding the mental processes in question. And if the simulation model predicts RTs that conform to the experimental data, this indicates that the proposed processing view is at least not incompatible with psychological reality.

This chapter first gives a general description of the model for each of the three task situations: bimodal vowel-detection (Chapter 4), auditory vowel-detection (Chapter 3) and modality decision (Chapter 5). In building up the model I take into account theoretical considerations (e.g., presumed stages of processing) and empirical constraints (e.g., obtained single-channel means). Next, I present an implementation of the model based on Vorberg's (1989) approach of mathematical models involving parallel and interactive channels. This implementation uses as few parameters as possible to account for the data of the bimodal vowel-detection experiments (reported in Chapter 4). After presenting the goodness of fit of this relatively simple model with the data, the effects of certain parameter variations are examined, as well as the ease with which the model can be extended to account for the data of the auditory vowel-detection experiments. The results obtained with the modality decision task are not simulated, because I feel more research with that task is necessary. Third and finally, the merits, problems and possible developments of the implemented model are discussed.

7.1 A cognitive view on processing

Cognitive psychology in general and psycholinguistics in particular consider processing to consist of a series of computations performed on internal representations (cf. Chapter 1). As Bower (1975) stated:

> "The information-processing approach assumes that perception and learning can be analyzed conceptually into a series of stages during which particular components ... perform certain transformations or recodings of the information coming into them. The subject's eventual response ... is considered to be the outcome of this lengthy series of operations. Each stage in the system receives as input the information as coded in its predecessor stage, operates upon it so as to condense, abstract, recode, or elaborate it, and then passes this product along to the next stage in the analysis. Since external stimuli cannot get inside an organism, the representation of them ... and their interactions ... is what we call "information", and this is the content we describe in our theories" (Bower, 1975, p. 33).

Figure 7.1 illustrates this cognitive view of processing for bimodal stimulus presentation. The presumed stages of visual or auditory processing, during which particular important transformations in the representational format of information occur, are represented by boxes (see Kosslyn, 1981, for a description of different types of such transformations). The lines connecting the boxes or stages represent the transfer of information to later operating stages. It should be noted that a box in the diagram does not correspond to a "module": Input to and output from boxes represent (complex) system states at particular moments in time, but which information led to the occurrence of those states, or where it came from, is left completely unspecified. The processing states are sequential, but the information flow within the system can still be interactive. By indicating changes in system states over time, Figure 7.1 incorporates temporal aspects of processing. Structural aspects follow from the limited ways information can be transformed over time or exchanged between representational units in the system (not visible in the figure).

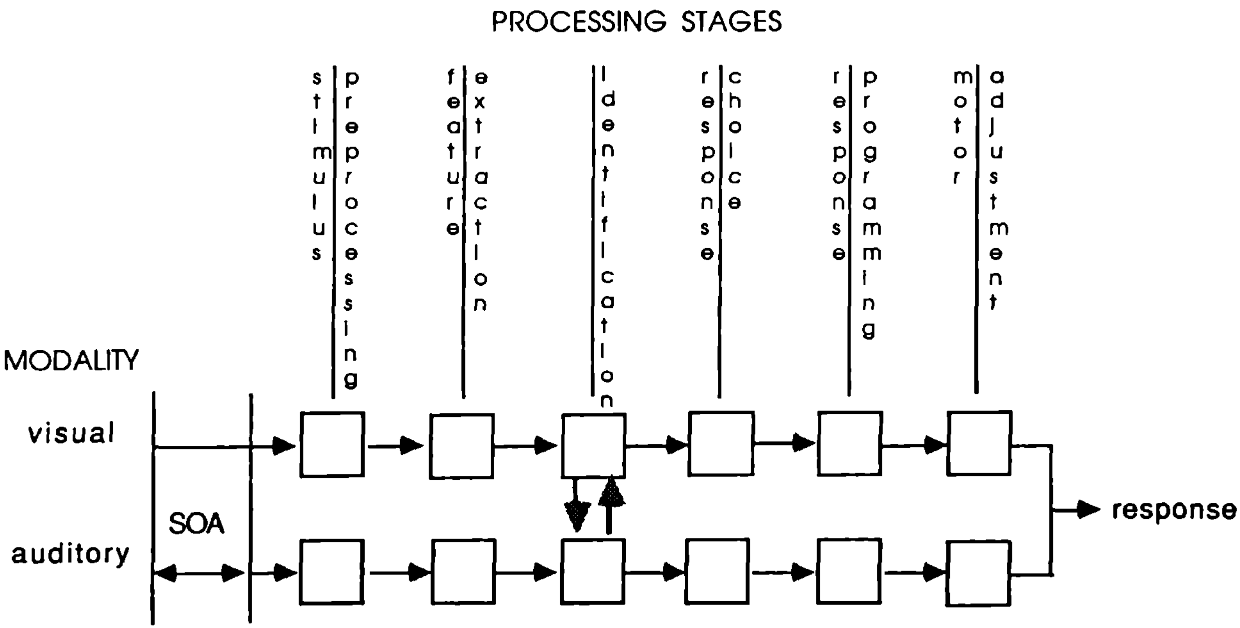


Figure 7.1. A model for bimodal processing.

7.2 A model for bimodal sublexical processing

7.2.1 General considerations

A useful and empirically well-supported framework for a bimodal processing model is provided by Sanders (1980, 1983, 1988). In his "cognitive-energetic model of arousal, stress and performance" information to be processed passes through six (in some publications four) serial stages between presentation and response (I will not consider the energetic level of the model here). These stages, which were established by way of additive factor studies (Sternberg, 1969), are the following: stimulus preprocessing, feature extraction, identification, response choice, response programming, and motor adjustment. For the bimodal situation I also assume that visual and auditory processing both pass through these six serial stages, but I add the possibility of parallel processing in both modalities. The onset time of processing in each channel is determined by the SOA.

Ignoring for a moment cross-modal interactions, the model depicted in Figure 7.1 can be considered the independent *race model* that played such an important role in the previous two chapters. For the bimodal vowel-detection task of Chapter 4, the RT predicted for a bimodal signal would be the minimum time necessary to pass through all processing stages in either the visual or the auditory channel. For the modality decision task of Chapter 5, the RT to the visual channel would be the time necessary to pass through all visual stages, but only if for that particular trial the visual stimulus was faster than the auditory one (analogously for the reactions to the auditory modality).

Application of the race model to the auditory vowel-detection task of Chapter 3 would lead to a predicted RT for bimodal stimuli that is the sum of the processing times for all stages in the auditory modality, which amounts to the auditory single-channel RT. Since the RTs to bimodal stimuli were facilitated relative to the single-channel RTs in those experiments, it is clear that the model developed here must include various types of *interactions* between the modalities.

Interactions between the visual and auditory channels may prolong or shorten the duration of processing stages in each of these. To simulate the just mentioned facilitation of the bimodal conditions in the auditory vowel-detection task (ascribed to preparation enhancement), any visual stimulus shortens the duration of processing of the auditory target proportional to the time it precedes the target. In other words, the larger the time lag or SOA between the visual and the auditory stimulus, the larger the facilitation effect, at least within the SOA-range that was used in the experiments.

More important facilitatory interaction effects occur at the identification level in congruent conditions (such as Aa) in all experiments. As soon as a stimulus in a channel is processed up to the identification stage, a representational unit is assumed to become and stay active until a response is given. An active unit can reduce the duration of the identification stage for a stimulus in a second channel, as far as processing of this second stimulus has not yet passed that identification stage. I take the reduction in processing time to be *proportional to the time that the second channel can be influenced* by the first. For example, if the identification stage in the second channel is reached after that in the first, it will be shortened more (in ms) when it takes a long time to complete than when it is quickly passed through. Furthermore, if processing of the second stimulus is already in the identification stage when the first stimulus reaches its own identification stage, only the remaining part of the stage will be shortened.

For the two-choice response situation in the auditory detection task (Experiment 3, Chapter 3), inhibitory effects arise in the vowel-incongruent conditions (e.g., Ae) from a similar but negative (i.e. processing time increasing) interaction of the response choice stages in the two channels (thus, they are *not* symmetric to facilitatory effects). The inhibitory effects found in Experiment 3 were ascribed to response competition, and were not obtained in the go/no-go bimodal detection task. For the incongruent conditions (e.g., Au) in Experiments 4, 5, and 6, I simply assume that the prediction by the race model suffices.

It will be clear from this account that I assume differences in processing between bimodal congruent, incongruent and neutral conditions in the three types of tasks, depending on specific characteristics of the tasks. In Chapter 6 it was suggested that in addition the allocation of attention to the visual and auditory modalities differs for the focused and divided attention variants: Attentional capacity may be divided more evenly over both modalities in the divided attention task than in the focused attention task. However, I will examine here how well the model predicts both types of results without representing this difference in attention allocation.

7.2.2 Specific implementation assumptions of the model

1. Processing involves six serial stages in both the visual and the auditory channel. Within each channel, the processing durations of the successive stages are considered to be independent random variables. If there are

no intermodal interactions, processing time for a stage i (where i=1, .., 6), follows a gamma-distribution with parameters λ_i (= rate) and m_i (= number of sequential exponentials to generate the gamma-distribution in question). As a simplifying assumption, for all stages within a channel the (unmodified) processing rates are equal: $\lambda_i = \lambda$. The total processing time within a channel for the six stages combined will now be distributed according to a gamma-distribution with parameters λ and n, where n = $m_1 + ... + m_6$. The mean total processing time will be μ_n=n/λ, the mean duration of each substage i will be μ_i=m_i/λ.

Use of the gamma-distribution is motivated by the following considerations. First, gamma-distributions with n=3 or more fit the usually right-skewed RT-distributions very well (McGill, 1963). While the exponential distribution often arises, in practice, as the distribution of the amount of time until some specific event occurs, the gamma distribution arises as the distribution of the amount of time one has to wait until a number of events has occurred (Ross, 1988). Second, the variance of a gamma-distribution has a reciprocal relation to parameter n (keeping the mean constant, it decreases when n increases). This is important to us, since the amount of statistical facilitation observed in the bimodal redundant conditions of the divided attention task depends on the variance of the contributing visual and auditory components (the larger their variance, the smaller the mean RT of their minimum distribution). Treating the n's for the visual and auditory channels (n_v and n_a) as free parameters, we should be able to find those values of them that result in the right amount of statistical facilitation and in a good fit between the model and the empirical data.

2. To simulate cross-modal activation effects during certain processing stages, estimates of the onset and duration of these stages are required.

To obtain such estimates, several assumptions must be made. For example, if the stages are indeed serially organized (which is by no means undisputed), estimation is only possible on the basis of a comparison of results obtained with different tasks, and thus assumptions about their interrelation are needed. An important assumption, often implicitly made in psychological research, is that of "pure insertion" (Donders, 1869/1969): adding or leaving out a stage does not affect the duration of the processing stages the tasks have in common. However, the validity of this assumption has been criticized (e.g., Sternberg, 1969). As an example of how the "pure insertion" assumption could lead to some estimation of stage durations, consider the following. In the bimodal vowel-detection task I obtained mean

RTs of about 380 ms, implying that processing passes all stages within 380 ms. Simple RT-tasks in which a visual or auditory signal must be detected lead to a lower bound of 120 ms (cf. McGill, 1963). If such tasks required only preprocessing, response programming and motor adjustment, the "pure insertion" assumption would lead to the conclusion that the other stages together (feature extraction, identification and response choice) take about 380-120=260 ms. The experimental literature offers practically no clues about the average duration of each stage (which, of course, may be task and subject dependent), but it seems that peripheral processing is reasonably fast, while the identification and response choice stage may take more time and are more variable in duration (cf. Schmidt, 1988, pp. 64-66, 177, 235). Tentatively, the following estimates of the durations of different stages could be given.

1 - preprocessing: 20-50 ms
2 - feature extraction: 20-50 ms
3 - identification: 100-150 ms
4 - response choice: 100-150 ms
5 - response programming: 20-50 ms
6 - motor adaptation: 20-50 ms

For most simulations, parameters m_1, m_2, m_3, and m_4 were set at fixed values that led to average durations of the associated stages within the just given ranges. Table 7.1 shows the values of a characteristic set of m-parameters for different stages in the visual and auditory modality, and the resulting duration of the stages (computed by means of the equation $\mu_i = m_i/\lambda$), when a total processing time of 380 ms is assumed, n_v=25, and n_a=22. Parameters n_v and n_a (which were equal to the sum of the m's for the modality in question) were varied in the simulations via the combination of m_5 and m_6 (since there were no interactions in stage 5 and 6, it was not necessary to differentiate between them in the simulations).

NR	NAME	M_i	VISUAL DURATION	M_i	AUDITORY DURATION
1	preprocessing	2	(30.4 ms)	2	(34.6 ms)
2	feature extraction	2	(30.4 ms)	2	(34.6 ms)
3	identification	10	(152.0 ms)	8	(138.4 ms)
4	response choice	7	(106.4 ms)	6	(103.8 ms)
5	response programming				
			varied		varied
6	motor adjustment				

Table 7.1. Values of the m-parameters for each stage used in most simulations, and the resulting average duration of each stage when a total processing time of 380 ms is assumed and the n-parameters are set at 25 (visual) and 22 (auditory). In this example, $m_5+m_6=4$ for both the visual and the auditory modality (total duration of stages 5 and 6 combined: 60.8 ms for the visual modality, 69.2 ms for the auditory modality).

3. We now turn to some implementation assumptions concerning specific conditions and aspects of the experiments. For congruent conditions in all experiments, cross-modal activation can occur as soon as the identification stage in a channel is reached: When processing in the other channel next passes through the identification stage, processing time is reduced to an extent determined by parameter α. An example will make this more concrete. Suppose the total processing time μ_n=380 ms and n=25. Given the relation $\mu_n=n/\lambda$, λ will be equal to 1/15.2. If $m_3=10$, the duration of the identification stage is $m_3/\lambda=152$ ms (cf. Table 7.1). Given that the cross-modal activation parameter $\alpha=.50$, its duration will become $(m_3/\lambda)\alpha=76$ ms. Processing in a channel is only facilitated by cross-modal activation during the identification stage. The same α is applied to cross-modal influence from visual to auditory and vice versa.

4. For the incongruent conditions in the auditory vowel-detection task, cross-modal inhibition can occur as soon as the response choice stage in a channel is reached: When processing in the other channel then passes through the response choice stage, processing time is increased by a certain amount: Instead of (m_{j}/λ) ms, its duration becomes $(m_{\text{j}}/\lambda)\iota$ ms, where ι stands for the cross-modal inhibition parameter. For example, if the response choice stage in isolation takes 120 ms, it would take 180 ms under the influence of inhibition when ι=1.50. Processing in a channel is only inhibited cross-modally during the response choice stage. The same ι accounts for the cross-modal influence from visual to auditory and in the other direction.

5. In this first exploration, no interaction is assumed to occur in the incongruent conditions of the bimodal vowel-detection task (though the obtained coactivation effects for these conditions in Experiments 4, 5 and 6 indicate the situation must be more complicated than just a race).

6. Different grapheme-phoneme combinations do not require changes in the interaction parameters of the model. Differences in single-channel RTs between different letters and speech sounds are simply divided equally over all n substages. In other words, if the mean RT increases, this is not due to a change in n, but in λ, the general processing rate. (An alternative to be explored is to restrict differences in processing duration to the first three processing stages, assuming that response stages are equal for different stimuli).

7. Stimulus onset asynchrony is modeled by simply having processing in one channel start a certain period of time before processing in the other channel. The expected processing time for a delayed signal, measured from the onset of the first signal, increases by SOA ms. The model must account for the RTs obtained not only in different conditions, but also at different SOAs.

8. As far as they are applicable, parameters are assumed invariant across conditions and experiments. This is a strong assumption, since it implies, for example, that the same set of parameter values can be used for both the blocked and the mixed variants of the bimodal vowel-detection experiments (Chapter 4). Finally, in this first exploration I simulate the mean RTs over all subjects, taking all assumptions made in the model to be valid for the group RT-distribution (see Appendix 1 for consequences of

this choice).

7.3 Results of parameter estimation and model fitting

To simulate the RT in a bimodal redundant condition, the model receives the single-channel RTs for the letters and speech sounds constituting that bimodal condition. When the response time to a single channel is gamma-distributed, the mean μ_n, known to be equal to n/λ, can be estimated by the mean RT for that channel. Thus, when n assumes a specific value, λ can be estimated ($\hat{\lambda}$ =n/RT). Given specific n's for the visual and auditory channels (n_v and n_a), the model computes the corresponding $\hat{\lambda}$'s and subsequently simulates the processing in a bimodal trial as follows. First, the durations of visual and auditory processing are determined for this particular trial, sampling randomly from gamma-distributions with the just described parameters n and $\hat{\lambda}$. Next, proceeding stage by stage, the bimodal process is followed and, depending on condition and task, facilitatory or inhibitory influences are exerted (subdivision of n in m_1, m_2, etc., determines when interactions should occur and how long they last). Depending on the task, the RT is now equal to the first channel to finish (bimodal detection), or the auditory processing time (auditory detection). For each combination of n_v and n_a the computation of the bimodal RT is repeated a number of times (from several hundred times for a small simulation to 35,000 times for a large one), in order to reduce the variance in the predictions caused by the random sampling (cf. Appendix 1). The procedure just described is called "Monte Carlo simulation" (see Maki & Thompson, 1973).

7.3.1 Bimodal vowel-detection task (divided attention)

Parameters were first estimated for the bimodal vowel-detection task (Chapter 4), because the largest number of RTs was collected for that task. The simulation results for the incongruent redundant conditions are considered first, followed by those for the congruent conditions. I assumed that the *incongruent conditions* need only two free parameters, namely the n's for the visual and auditory modality (other parameters, such as that for cross-modal activation, α, are set to 1, i.e. have no effect). In order to find the values of these parameters that result in the best fit between model and data, simulations were run with the sample size for the model equal to the number of replications in the experiment (see Appendix 1). Two different

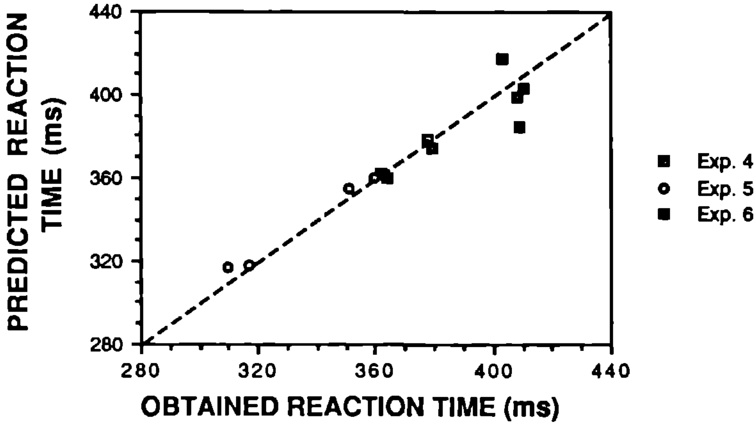


Figure 7.2. Obtained and predicted mean RTs for the incongruent conditions in the bimodal vowel-detection Experiments 4, 5, and 6. Parameter values: n_v=25, n_a=22.

indicators of the goodness of fit were computed. First, I computed the average absolute RT-difference between obtained and predicted RTs. Second, I computed a χ^2 statistic, taking into account the variance in the data and in the simulations (the formula for this statistic is given and derived in Appendix 1). For statistically significant χ^2s the model should be rejected, as a larger value of the χ^2 corresponds to a worse fit between data and model. Both the average absolute RT-difference and the χ^2 were computed including and excluding the data-points in Experiment 6 for SOA=100, since these points were considered unreliable (see Chapters 4 and 5 for a discussion of this issue). Furthermore, since different simulations resulted in different outcomes, the minimum and maximum values of both measures over 20 runs per data point were computed.

A whole range of n_v and n_a values was investigated by systematically varying the values of n_v and n_a from 15 to 35. For a large part of this range, the average absolute RT-difference between data and model was quite small (a difference of less than 8 ms was easily obtained). For the best combination of values for n_v and n_a found on both measures (n_v=25, n_a=22), Table 7.2 gives some information on the variation in RT-differences between data and model. It indicates, for example, that for the 20 runs with a sample size equal to the experimental number of replications, the RT-differences between data and model for the incongruent conditions varied between 3.6 (MIN) and 6.4 ms (MAX), when the two data points for SOA=100 (Experiment 6) were excluded (keeping 10 data points). For this combination of n_v and n_a values, I ran an extra simulation with 35,000 samples per data point. Figure 7.2 graphically illustrates the results of this large simulation for the datapoints of all incongruent conditions in Experiments 4, 5, and 6 (see Appendix 2 for the exact obtained and predicted RT-values). The large simulation resulted in an average absolute RT-difference of 3.76 ms. Though several of the runs with smaller sample size resulted in insignificant χ^2s, the large simulation led to a significant χ^2 (for 10-2=8 degrees of freedom: χ^2=27.23, $p<.01$).

	INCONGRUENT		CONGRUENT		ALL	
	MIN-MAX	(NR)	MIN-MAX	(NR)	MIN-MAX	(NR)
Diff.	3.6–6.4	(10)	3.5–6.3	(7)	3.5–6.4	(17)
	5.2–9.2	(12)	4.0–7.7	(9)	4.7–8.6	(21)
	MIN-MAX	(NR)	MIN-MAX	(NR)	MIN-MAX	(NR)
χ^2	13.0–52.7	(10)	11.2–31.6	(7)	24.2–84.3	(17)
	19.2–93.4	(12)	12.4–65.0	(9)	31.7–158.4	(21)

Table 7.2. Statistics of the simulation results for incongruent and congruent redundant conditions for parameter values n_v=25, n_a=22, and α=.76. Diff: average absolute RT-difference between model and data. NR: number of datapoints in the simulation (excluding or including Experiment 6).

Table 7.2 also shows the range of the results for the modeling of the *congruent conditions*. Taking n_v=25 and n_a=22 as in the incongruent conditions and determining the best value of parameter α to account for cross-modal activation at the identification level, I simulated the RTs, again with and without the two data points for SOA=100 from Experiment 6. Without these data points, the average absolute difference between obtained

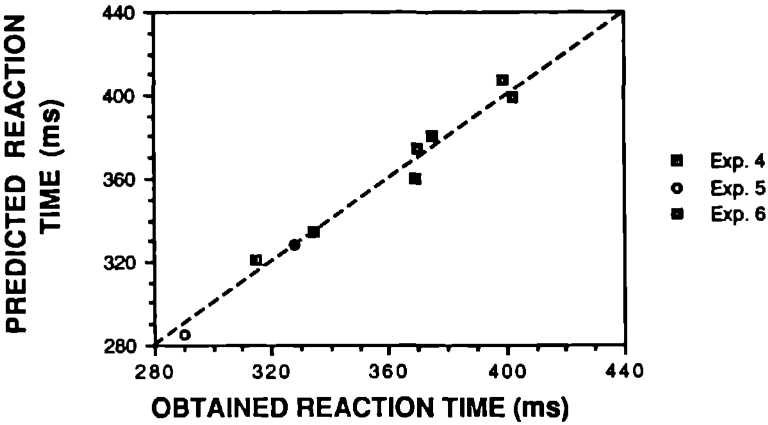


Figure 7.3. Obtained and predicted mean RTs for the congruent conditions in the bimodal vowel-detection Experiments 4, 5, and 6. Parameter values: n_v=25, n_a=22, α=.76.

and predicted RTs (obtained in 20 runs with sample size equal to number of experimental replications) varied between 3.5 and 6.3 ms for α=.76. Again, for the best runs the χ^2s (with 7-1=6 degrees of freedom) were not significant. A large simulation with 35,000 samples resulted in an average absolute RT-differences between data and model of 4.01 ms, but led to a significant χ^2 of 31.65 ($p<.01$). The mean RTs for the congruent conditions that were predicted by this large simulation are graphically represented in Figure 7.3; their exact values can be found in Appendix 2.

As was indicated in Table 7.1, m_1, m_2, m_3, and m_4 were set at fixed values. Thus, the duration and onset time of the four corresponding stages (among which the identification stage) were not directly manipulated in the simulations just described. Since for the congruent conditions cross-modal activation effects operate only at the identification level, it is interesting to investigate to which extent the assumed length of the identification stage and its moment of onset influence the predictions of the model. The duration of the identification stage can be changed by varying the value of m_3, its onset by varying m_1 and/or m_2. To find the best fit for one value setting, the cross-modal activation parameter α must, of course, be re-adjusted with each change. When the duration of the identification stage was changed more than about 50 ms, this change could not fully be compensated via adjustment of α anymore (resulting in a worse fit). However, for most onsets of the identification stage (early or later) in processing quite good fits were found. Thus, an exact positioning of the identification stage in processing was not critical to obtain reasonable predictions. Both of these findings may be related to my use of only two (or three) SOAs for the simulations: RTs for more SOAs might have required more precise settings of the duration and position of the identification stage. Since no exact information is available concerning the temporal locus of the identification stage in bimodal processing, the flexibility of the model could be seen as an advantage, as it can easily adapt to future results that provide more definite knowledge concerning this locus; but such flexibility of course also makes the model harder to falsify.

Finally, Table 7.2 gives the average absolute RT-difference and χ^2 for the *combination of congruent and incongruent conditions* (ALL). Since I first analyzed the incongruent conditions, and subsequently added the congruent conditions, the values of the parameters chosen did not necessarily lead to the best over-all fit. However, instead of further investigating this issue, I thought it more interesting for a first exploration to see how well the just obtained values could be used to simulate the obtained RTs in the auditory vowel-detection experiment (Experiment 3).

7.3.2 Auditory vowel-detection task (focused attention)

Several problems arose when I tried to apply the model to the vowel-conditions of Experiment 3 in Chapter 3 (auditory vowel-detection). First, only RTs to the auditory stimulus were available: Thus, RT-values for the visual stimuli A and E had to be estimated. The results of the bimodal vowel-detection experiments indicated that these RTs may vary a lot, but on the average they were about equal to their auditory counterparts. In this exploration I therefore simply took as input to the model equal visual and auditory single-channel RTs.

The auditory single-channel RTs were the source of a second problem, since the RTs for /a:/ and /e:/ turned out to be practically identical (423 ms). If the input to the model were identical for all conditions involving "A" and/or "E", then of course the predicted outcome would be identical. The data, however, showed large RT-differences between bimodal condi-

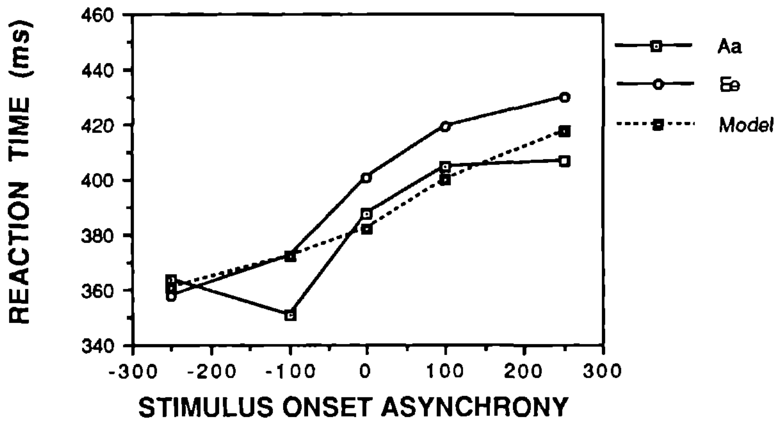


Figure 7.4. Obtained and predicted results for the congruent conditions in the auditory vowel-detection task over five SOAs.

tions involving auditory /a:/ and /e:/, conditions with /a:/ being 15 ms or more faster than those with /e:/. The simulation model could therefore never produce a good fit with both types of conditions. To get a general idea of how well the model fits despite this problem, I present the model's predictions next to the results obtained for both "A" and "E", renouncing the computation of the average absolute RT-difference and the χ^2.

A third problem was that the results for the bimodal neutral conditions in the auditory vowel-detection task showed a general influence of the visual modality on the auditory (see Chapter 3). The extent to which auditory processing was facilitated by a visual stimulus was accounted for by a parameter δ, which exerted an SOA-dependent influence (the more the visual stimulus preceded the auditory, the larger the cross-modal influence became).

Keeping n_v, n_a and α at the values that have been determined above, and varying δ, I predicted the RTs for the *vowel-congruent conditions.* Figure 7.4 illustrates the results for a run with 35,000 samples with δ set at .91. The obtained and predicted RTs fall within the same range, given

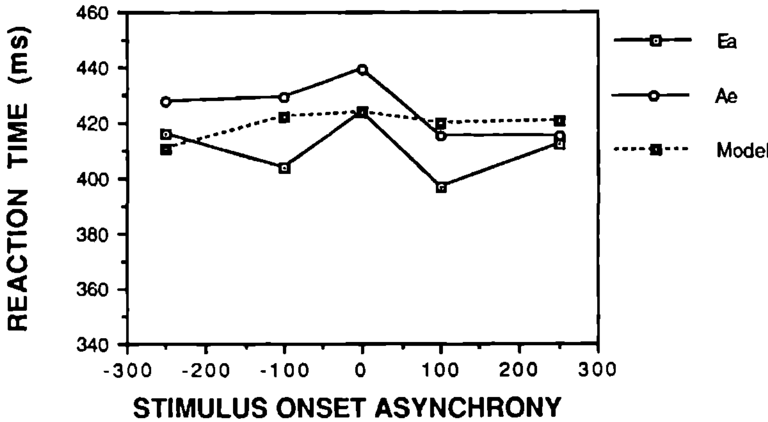


Figure 7.5. Obtained and predicted results for the incongruent conditions in the auditory vowel-detection task over five SOAs.

equal visual and auditory single-channel RTs as input (see Appendix 2 for the exact values). I also simulated the RTs for the *vowel-incongruent conditions*, using as parameters n_v, n_a, δ and an inhibition parameter ι (to stand for the cross-modal inhibition during the response choice stage). For n_v, n_a and δ set as before, and ι equal to 1.21, the results depicted in Figure 7.5 indicate that, again, the model produced RTs in the same range as the obtained data. An interesting test of how reasonable δ is, is to set α=1 and compare the results with only parameter δ included to those of the *bimodal neutral conditions* (e.g., Pa or *e). As Figure 7.6 shows, the resulting RT-pattern was quite similar to that observed in the data.

7.4 An evaluation of the simulation model

In this section, I first sketch some general characteristics of the implemented model. Next, I evaluate the model as far as it is implemented, by considering the model's fit to the data of the bimodal and the auditory vowel-detection experiments. Finally, I give suggestions for improvement of the model.

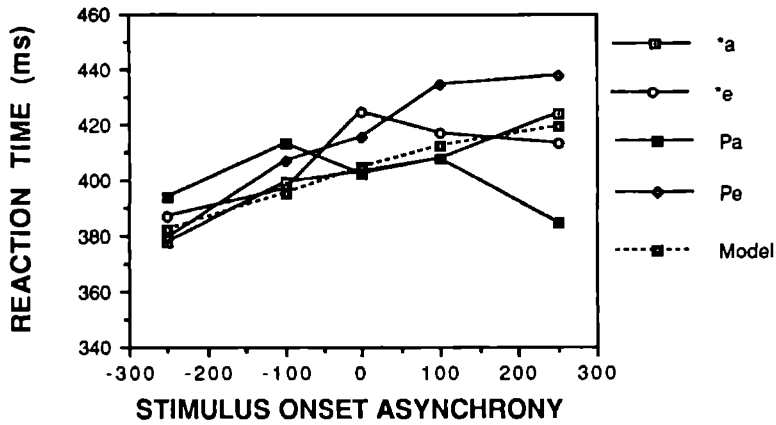


Figure 7.6. Obtained and predicted results for the neutral conditions in the auditory vowel-detection task over five SOAs.

The implemented model expresses about the most simple view of bimodal processing one can think of: Processing proceeds in both modalities in parallel, but serially in each. In specific theoretically motivated stages, processing in one modality may influence processing in the other modality. The temporal duration of each stage is not constant, but varies to some extent from trial to trial. Though I have considered only the mean RTs predicted by the model, it also predicts variances, and, in fact, the whole distribution of the RTs. The intrinsic stochastic nature of the model gives it an advantage over deterministic types of models, since such models need to add some stochastic process (e.g., during responding) to account for variability in and between subjects' performance. Also, the model needs no transformation rule to turn model-output into RT (connectionist models, for example, often need assumptions to relate "number of iterations" or "time cycles" to RT, cf. Phaf, 1986, p. 80-82; Phaf, van der Heijden, & Hudson, 1990).

The model does not make any specific claims about what exactly takes place within each substage of processing. Different conceptions of processing in a substage (e.g., with respect to feature extraction) are compatible with the model, as long as the time to complete the substage is gamma-distributed. A further specification of the different stages is perhaps possible on the basis of the experimental literature.

Related to this point is the fact that the simulation model does not have explicit representations for stimuli, and in a sense reflects only the influence of ongoing processes on other processes. However, assumptions concerning the time during which a certain influence remains active can be translated into parameters that operate for a certain time period or when certain processing stages are active. The model therefore has possibilities similar to models that assume explicit representations (cf. connectionist models).

How well did the model fit? In terms of the average absolute RT-difference between data and model, the model gives a satisfactory simulation of the data obtained in the *bimodal vowel-detection experiments*, even with only few parameters. The same small set of parameters seemed adequate to simulate results for different combinations of letters and speech sounds, presented at different SOAs and under diverse task situations.

However, most χ^2s computed for the simulations were significant, indicating that the model in its current form should be rejected. One reason for the relatively large χ^2s can be found in noise in the single-channel RTs that serve as input for the model. For example, the RTs for the single-channel conditions of Experiment 5 that were allocated to the SOA of -100 or 0 were different (e.g., 351 vs. 360 ms), even though they concerned identical tokens (cf. discussion of Experiment 5).

Another reason for the large χ^2 lies in the high quality and quantity of the obtained data. First, the data showed a rather small variance. Second, the number of replications for each empirical data-point was quite high (e.g., 50 replications for each of 35 subjects = 1750 measurements in Experiment 4). Both the data variance and the number of replications strongly influence the size of the χ^2 (see Appendix 1, s^2_{data} and n_t).

However, close inspection of the simulation results suggests that the large χ^2 is due in part to differences between experiments. Figures 7.2 and 7.3 and the associated Table in Appendix 2 show that a good fit for the incongruent conditions in Experiments 4 and 5 is accompanied by a relatively poor fit for those in Experiment 6. The RTs predicted by the model for Experiment 6 are generally faster than those obtained. It is interesting to note that a similar, but somewhat smaller effect was obtained with the race model of Chapter 4 (see Table 4.6). That race model was based on the single-channel distributions as a whole, while the model presented here assumed gamma-distributions with means equal to those of the single-channel conditions. These results indicate that further investigation should not be restricted to a mere comparison of obtained and predicted means, but should include a more detailed scrutiny of the obtained and predicted bimodal distributions.

It could be speculated that the observed difference in fit between the experiments results from subtle differences in task performance between the blocked and mixed variants of the bimodal vowel-detection task, e.g., in terms of attention allocation or decision processes. It may be noted in this context that the number of possible targets (in other words, the size of the memory set) was twice as large in Experiment 6 as in Experiments 4 and 5 (two visual and two auditory targets vs. one visual and one auditory target). Sternberg (1969) has demonstrated that this factor affects response time in high-speed memory scanning tasks, in which subjects decide as fast as possible whether or not a probe stimulus belongs to a small memorized set of items. Further research should clarify whether this factor also played a role in my detection experiments or not. However, in generating this hypothesis the simulation model already proves to have some heuristic value.

The results that were obtained with the same settings of the parameters for the *auditory vowel-detection experiment* (Experiment 3) are encouraging. The predicted RTs fall within the same range as the obtained. Furthermore, the choice of visual RTs equal to the obtained auditory RTs was quite arbitrary. Representing processing differences between letters (A, E) and speech sounds (/a:/, /e:/) by two extra parameters should lead to a much better fit. Also, I have not considered differences in attention allocation between the divided and focused attention tasks.

The simulation for the auditory vowel-detection task assumed a general influence of a visual stimulus on the auditory signal, the size of which was determined by a parameter δ. Further research should clarify if a similar influence was present in the bimodal detection task.

If new parameters are added to the model, the use of an optimization routine (e.g., Davidson-Fletcher-Powell, see Levelt, Schriefers, Vorberg, Meyer, Pechmann, & Havinga, submitted) becomes indispensable. Even for the relatively simple model presented here, such a routine is potentially very helpful. However, application of such a routine to the model led to complications. Because of the stochastic nature of the input to the model, the predicted RTs varied from simulation to simulation. This resulted in repeated failures of the optimization routine. Increasing the sample size to reduce the variance led to unacceptably long run times. A solution to this problem would be to transform the model reported here into an analytic model, that mathematically derives the exact expected bimodal RT-distributions. The feasibility of this approach for similar situations has been demonstrated by Vorberg (in preparation).

Finally, instead of using group mean RTs and variances (see Appendix 1), it would mathematically and psychologically be more correct to apply the model to the data of single subjects. Since the number of replications per subject for a particular condition is quite high, parameter estimation and model fitting should be possible for each subject separately. The resulting χ^2s can afterwards be added over subjects (adapting also the number of degrees of freedom).

To summarize, several changes and extensions may result in a better-fitting and psychologically more realistic model. However, the implemented version of the model presented here, which had only a very small number of parameters and theoretical constructs, was already capable of giving a reasonable fit to the empirical data in terms of the average absolute RT-difference. This result supports the general theoretical approach of bimodal sublexical processing expressed in Chapters 6 and 7, and indicates that further development of the model will surely be worthwhile.

Summary

Both the spoken and the written form of Dutch and English, as well as of other Western languages, use representation systems which involve modestly sized sets of basic symbols: about 40 phonemes in the spoken language mode, 26 letters in the written mode. This thesis was concerned with various aspects of the relationship between the mental representations of these visual and auditory symbols, and how they interact in linguistic processing.

A review of the experimental literature (Chapter 2) indicated that relatively little firm knowledge is available concerning the issue. Many problems remain unsolved. Can the mental representations of letters (graphemes) activate those of speech sounds (phonemes) during early stages of linguistic processing? And vice versa, can phonemes activate graphemes? How quickly can such activation take place, if it occurs at all? Can such cross-modal influence only be of a facilitatory kind, or can it also be inhibitory? Does it occur regardless of the subject's intention or conscious control, i.e. is it automatic?

Seven experiments were conducted to answer these questions. A first series of three experiments (reported in Chapter 3) investigated the structural and temporal relationship between graphemic and phonemic processing by means of a cross-modal priming procedure. In these *auditory vowel-detection* experiments, Dutch subjects made a forced choice on the identity of the vowel in an auditorily presented syllable (e.g., /a:/ or /e:/ in syllables such as /pa:/ or /ke:/). To determine if and when phonemic representations are activated by graphemes, visual letter primes (e.g., P, A or E) were presented before, during or after presentation of the syllable. In two experiments, the presented letter was either congruent with the consonant of the auditory CV-syllable (e.g., letter P with syllable /pa:/), or incongruent (e.g., K with /pa:/). Faster reaction times were obtained over a

broad range of stimulus onset asynchronies (SOAs) for congruent than for incongruent consonant-priming conditions. In a third experiment, the relationship between the presented letter and the target vowel itself was also varied. SOA-dependent facilitation effects were found with respect to a bimodal baseline-condition when the prime was congruent with the target vowel (e.g., A with /ka:/) and inhibition effects when it was congruent with the competing target vowel (e.g., E with /ka:/). Whereas the facilitation effects were interpreted as an indication of cross-modal activation effects at a representation level, the inhibition effects were considered to be the consequence of response competition. The results of the three experiments support the hypothesis of grapheme-to-phoneme activation between sub-lexical representations, and indicate that such activation is automatic in nature.

A second series of three experiments (reported in Chapter 4) was conducted in order to demonstrate phoneme-to-grapheme activation effects, and to examine if cross-modal representational inhibition effects may also occur. Involving a different experimental paradigm (go/no-go detection), these experiments were further intended to provide more detailed information concerning the temporal aspects of cross-modal activation. In three *bimodal vowel-detection* experiments, subjects were to detect visual and/or auditory vowel targets. Unimodal or bimodal stimuli were presented that, for the go-conditions, consisted of two targets (one visual and one auditory), one target combined with a neutral stimulus, or one target (either visual or auditory) presented in isolation. In some two-target conditions, the visual and auditory stimuli were nominally identical or name-congruent (e.g., visual A, auditory /a:/), in others they were not (e.g., U, /a:/). Temporal aspects of cross-modal activation were investigated again by varying the SOA of visual and auditory component stimuli.

If grapheme and phoneme representations co-activate each other, reaction times to name-congruent bimodal conditions should be faster than to name-incongruent conditions, after differences in the distributional characteristics of the visual and auditory components have been taken into account. It was first shown that coactivation did indeed exist by a comparison of the reaction times obtained and those predicted under maximal separate activation. Subsequently, the obtained reaction times were compared to those expected for a race model involving only independent separate activation. The expected curves were violated over a much longer temporal range and to a larger extent for the congruent than for the incongruent conditions.

The specific pattern of results indicated the existence of a fast bidirectional cross-modal facilitation spreading between grapheme and phoneme representations, and an absence of cross-modal inhibition effects at the representation level. Furthermore, general characteristics of visual and auditory stimuli seemed to lead to asymmetric patterns of results when SOA was varied.

In a seventh experiment (reported in Chapter 5), a third paradigm was used to extend the prediction method based on the hypothesis of a race to recognition between signals presented in the two modalities, and to test an alternative explanation of earlier results in terms of a dominating tendency to react to the visual modality ("visual dominance"). In a *modality decision* task, subjects were asked not only to detect specific visual and/or auditory target signals, but also to indicate the modality in which they first identified such a target (by pushing either a SEE or a HEAR response button). The results, analyzed in terms of both percentage and speed of reactions to a modality, in general confirmed the findings of earlier experiments: Larger facilitation effects were found in the congruent than in the incongruent two-target conditions with respect to predictions based on a race-model under the majority of seven SOAs, both to visual and to auditory targets.

On the basis of the seven experiments just described, a coherent view of cross-modal activation effects between graphemes and phonemes in bimodal sublexical processing could be developed (Chapter 6). According to this view, processing of letters and/or speech sounds leads to a fast and automatic spreading of activation of graphemes to associated phonemes and vice versa, but not to any cross-modal inhibition effects between representations. Furthermore, the effects of the three types of experimental tasks on the subjects' performance were considered, and the results were related to findings in the domain of word recognition.

Since the results of the present research involving sublexical material (syllables, single letters and speech sounds) are consistent with those obtained in word recognition research, this suggests that fast grapheme-phoneme effects also play a role there. Having established the validity of the present experimental paradigms, new experiments including both lexical and sublexical material can now be designed to answer more detailed questions concerning the relationship between cross-modal activation and word recognition.

In the last chapter of the thesis (Chapter 7), a model was presented that was intended to formalize the bimodal processing view developed in earlier chapters. In the model it was assumed that processing in the visual and the auditory modality passes through a number of successive stages, that each take a gamma-distributed period of time to complete. A further assumption was that activation is spread from one modality to the other as soon as an identification stage is reached, but only between congruent grapheme and phoneme representations (e.g., those for letter A and speech sound /a:/). No inhibition effects were supposed to occur between incongruent graphemes and phonemes (e.g., those for letter U and speech sound /a:/). Monte Carlo simulations for the three bimodal vowel-detection experiments led to predicted reaction times that fit the empirical data quite well. A preliminary investigation of the model's predictive power for the auditory vowel-detection task also gave encouraging results.

Samenvatting

De gesproken en de geschreven vorm van het Nederlands (en van de meeste andere westerse talen) maken beide gebruik van een representatiesysteem met een beperkt aantal basissymbolen: de gesproken taal kan worden weergegeven met behulp van zo'n 40 fonemen, terwijl voor de geschreven taal 26 letters worden aangewend.

In dit proefschrift werden een aantal aspecten onderzocht van de relatie tussen de mentale representaties van deze visuele en auditieve symbolen, waarbij werd ingegaan op hun onderlinge wisselwerking tijdens de verwerking van taal.

Een overzicht van de experimentele literatuur (Hoofdstuk 2) gaf aan dat met betrekking tot deze kwestie relatief weinig gedegen kennis beschikbaar is. Veel problemen bleven onopgelost, waaronder de volgende. Kunnen de mentale representaties van letters (grafemen) die van spraakklanken (fonemen) activeren tijdens vroege stadia van taalverwerking? En omgekeerd, kunnen fonemen grafemen activeren? Als dat zo is, hoe snel kan een dergelijke activatie dan optreden? Kan zo'n intermodale invloed enkel faciliterend werken of tevens inhiberend? En tenslotte, treedt een dergelijke invloed onafhankelijk van de intentie of bewuste controle van de proefpersoon op, d.w.z. is er sprake van een proces dat automatisch verloopt?

Om deze vragen te beantwoorden werd een zevental experimenten uitgevoerd. In een eerste reeks van drie experimenten (gerapporteerd in Hoofdstuk 3) werd door middel van een "intermodale priming procedure" onderzocht of grafemen en fonemen elkaar activeren tijdens de verwerking van sublexicaal materiaal, en hoe snel dat geschiedt. In deze *auditieve vocaaldetectie*-experimenten moesten de Nederlandse proefpersonen beslissen tot welke van twee eerder gespecificeerde categorieën (bijv. /a:/ of /e:/)

de klinker in een CV-syllabe (zoals /pa:/ of /ke:/) behoorde.

Vlak voor, tijdens of juist na de auditieve presentatie van de syllabe (m.a.w. onder verschillende "Stimulus Onset Asynchronies" of SOAs) werd op een computerscherm een letter ("prime", zoals P, A of E) gepresenteerd. De invloed van de aanwezigheid van die letter op de vocaaldetectie werd in de eerste twee experimenten onderzocht door de relatie tussen letter en syllabe-consonant te variëren: er was sprake van consonant-congruentie (bijv. letter P met syllabe /pa:/) of van consonant-incongruentie (bijv. letter K met syllabe /pa:/). Onder verschillende SOAs werden voor congruente condities snellere reactietijden (RTs) verkregen dan voor incongruente.

In het derde experiment werd de relatie tussen de aangeboden letter en de doelvocaal zelf gevarieerd. Ten opzichte van een conditie waarbij in plaats van een letter een ster werd gepresenteerd, traden SOA-afhankelijke facilitatie-effecten op in condities waarbij de letter (bijv. A) congruent was met de doelvocaal in de gepresenteerde syllabe (bijv. /a:/ in /ka:/), terwijl inhibitie-effecten werden geconstateerd indien de letter (bijv. E) congruent was met de doelvocaal die niet in de gepresenteerde syllabe voorkwam (bijv. /e:/ bij presentatie van /ka:/). De facilitatie-effecten werden opgevat als een aanwijzing voor het bestaan van intermodale activatie tussen grafemen en fonemen; de inhibitie-effecten werden geïnterpreteerd als het gevolg van responsecompetitie. De drie experimenten tesamen ondersteunden de hypothese dat fonemen worden geactiveerd door grafemen bij de verwerking van sublexicaal linguistisch materiaal, en ze gaven aan dat een dergelijke activatie automatisch van aard is.

Om het bestaan aan te tonen van activatie-effecten in de andere richting, nl. van fonemen naar grafemen, en om te onderzoeken of intermodale inhibitie-effecten ook op representatieniveau kunnen voorkomen, werd een tweede reeks van drie experimenten uitgevoerd (gerapporteerd in Hoofdstuk 4). Deze experimenten waren verder bedoeld om meer gedetailleerde informatie te verkrijgen over de temporele aspecten van intermodale activatie. In deze *bimodale vocaaldetectie*-experimenten werd aan proefpersonen gevraagd zo snel mogelijk op een knop te drukken wanneer van tevoren gespecificeerde visuele en/of auditieve doelsignalen (zoals letter A of spraakklank /u:/) werden aangeboden, maar bij afwezigheid van dergelijke stimuli geen reactie te geven. Een reactie was vereist indien twee doelstimuli (een visuele en een auditieve) werden aangeboden, of indien één doelstimulus werd gepresenteerd (visueel of auditief), al dan niet vergezeld van een neutrale stimulus in de andere modaliteit.

Als twee doelstimuli werden aangeboden waren de visuele en de auditieve stimuli naamcongruent (zoals visueel A met auditief /a:/), of juist naamincongruent (bijv. U met /a:/). De tijdsaspecten van intermodale activatie werden wederom onderzocht door het variëren van de SOA tussen de visuele en auditieve stimuli in de bimodale aanbieding. Indien grafeem- en foneemrepresentaties elkaar wederzijds activeren, zouden de reactietijden voor naamcongruente bimodale condities sneller moeten zijn dan voor naamincongruente condities, tenminste nadat verschillen in de distributionele eigenschappen van de visuele en auditieve componentstimuli zijn verdisconteerd.

Met behulp van de enkele-kanaalsreactietijden kon worden voorspeld wat de snelst mogelijke reactietijden waren in bimodale condities met twee doelstimuli, wanneer werd uitgegaan van het ontbreken van intermodale activatie. Omdat de verkregen reactietijden sneller waren dan die welke voorspeld werden onder deze omstandigheden van "maximale gescheiden activatie", werd aannemelijk dat wederzijdse intermodale activatie daadwerkelijk was opgetreden. Vervolgens werden de verkregen reactietijden vergeleken met de voorspellingen van een racemodel, waarin geen afhankelijkheid tussen beide kanalen bestaat en waarin díe componentstimulus (visueel of auditief) van een bimodale aanbieding tot een reactie leidt, welke het eerst wordt geïdentificeerd (die m.a.w. de race naar herkenning wint). De voorspelde RT-curves werden geschonden over een veel langer tijdsbestek en in veel grotere mate voor de naamcongruente condities dan voor de naamincongruente condities.

Het specifieke reactietijdpatroon vormde evidentie voor het bestaan van een snel-optredende intermodale facilitatie tussen grafeem- en foneemrepresentaties, en pleitte tegen het bestaan van intermodale inhibitie-effecten tussen zulke representaties. Bovendien leken algemene eigenschappen van visuele en auditieve stimuli te leiden tot asymmetrische RT-patronen wanneer het relatieve aanvangsmoment (SOA) van visuele en auditieve doelstimuli werd gevarieerd.

In een laatste experiment (gerapporteerd in Hoofdstuk 5), werd de zojuist besproken voorspellingsmethode uitgebreid. In dit experiment werd nogmaals getracht intermodale activatie aan te tonen, maar tevens werd een alternatieve verklaring getoetst voor eerdere resultaten, waarbij een voorkeur van de proefpersoon voor reacties op de visuele stimulus werd aangenomen (in de literatuur wel aangeduid als "visuele dominantie").

In een *modaliteitbepalings*-taak moesten proefpersonen niet alleen vooraf gespecificeerde visuele en/of auditieve doelstimuli ontdekken, maar ook aangeven in welke modaliteit ze deze het eerst identificeerden (door op een ZIEN- of HOREN-antwoordknop te drukken). De resultaten van dit experiment, die werden geanalyseerd in termen van zowel het percentage als de snelheid van reacties op een modaliteit, bevestigden in grote lijnen de bevindingen van eerdere experimenten: rekening houdend met de voorspellingen op grond van een racemodel werden in de congruente condities met twee doelstimuli grotere facilitatie-effecten gevonden dan in de incongruente condities. Dit gold voor de meerderheid van zeven SOAs, en zowel voor visuele als voor auditieve doelstimuli.

Op basis van de zojuist beschreven zeven experimenten kon een samenhangende visie worden ontwikkeld ten aanzien van intermodale activatie-effecten tussen grafemen en fonemen bij bimodale sublexicale taalverwerking (Hoofdstuk 6). Volgens deze opvatting leidt de verwerking van letters en/of spraakklanken tot een snelle en automatische verspreiding van activatie van grafemen naar verwante fonemen en andersom, maar niet tot intermodale inhibitie-effecten. Verder werd de invloed van de drie typen experimentele taken op de reacties van de proefpersonen in de beschouwing betrokken, en werden de experimentele resultaten gerelateerd aan bevindingen in het woordherkenningsonderzoek.

Omdat de resultaten van het onderhavige onderzoek met sublexicaal materiaal (syllaben, enkele letters en spraakklanken) consistent zijn met die verkregen in woordherkenningsonderzoek, lijkt het aannemelijk dat snelle activatie-effecten tussen grafemen en fonemen ook tijdens de woordherkenning een rol zullen spelen. Nu de validiteit van de gehanteerde experimentele paradigma's is vastgesteld, kunnen nieuwe experimenten worden ontworpen met zowel lexicaal als sublexicaal materiaal, om een antwoord te krijgen op meer gedetailleerde vragen aangaande de relatie tussen intermodale activatie en woordherkenning.

In het laatste hoofdstuk van dit proefschrift (Hoofdstuk 7) werd een model gepresenteerd dat de bedoeling had de visie op bimodale taalverwerking die in de vorige hoofdstukken werd ontwikkeld, te formaliseren. In dit model werd aangenomen dat visuele en auditieve verwerking in een aantal opeenvolgende stadia plaatsvindt, die elk een gamma-verdeelde tijd nodig hebben om te worden afgerond. Een verdere aanname was dat activatie zich van de ene modaliteit naar de andere verspreidt zodra een identificatiestadium wordt bereikt, maar alleen tussen congruente grafeem- en foneemrepresentaties (zoals die voor de letter A en spraakklank /a:/). Er zouden geen inhibitie-effecten optreden tussen incongruente grafemen en fonemen (bijv. tussen U en /a:/). Monte Carlo-simulaties voor de drie bimodale vocaaldetectie-experimenten leidden tot voorspelde reactietijden die goed overeenkwamen met de empirische gegevens. Een eerste onderzoek naar de voorspellende waarde van het model voor de auditieve vocaaldetectie-taak leverde eveneens bemoedigende resultaten op.

Appendix 1: Aspects of the experimental analyses

The goal of this appendix is to explain in more detail some of the data-manipulation techniques and formulas of Chapters 4, 5 and 6. First a description is given of a method to average RT-distributions over subjects. This is followed by an examination of the formulas for independent and dependent statistical facilitation effects used in Chapter 4. Subsequently, the formulas are derived that were applied in Chapter 5 to compute the expected visual and auditory RTs and the proportions of visual and auditory reactions in the modality decision task. The appendix ends with a discussion of the statistical measures used in Chapter 6 to fit the simulation model to the empirical data.

A1.1 Reaction time distributions

For a given experimental condition or subject, one can construct a relative frequency distribution such as the one shown in Figure A.1 (adapted from Ratcliff, 1979). This distribution is computed by simply dividing the number of RTs with(in) a certain value (range) by the total number of reactions sampled. A relative frequency distribution is the empirical counterpart of a theoretical probability density distribution. It is often very useful (see Luce, 1986, and Townsend & Ashby, 1983) to represent the distribution of collected RTs cumulatively by computing the total relative frequency for all RTs smaller than or equal to a certain value (see Figures 4.1, 4.2 and 4.3 in Chapter 4). Using this "cumulative" or "empirical distribution function" (abbreviated as CDF in Chapter 4), one can then compute the RT associated with any given cumulative relative frequency and vice versa. The median of the RT-distribution can be easily observed (it is the RT corresponding to the cumulative relative frequency value of .50), and the steepness of the curve gives an indication of the variance (the

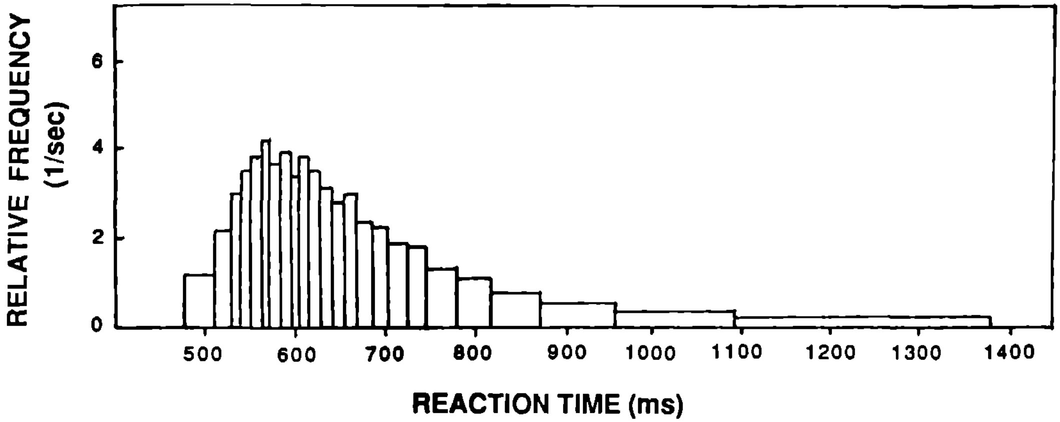


Figure A.1. Example of a relative frequency (group) distribution. Each bar represents a 2%-quartile.

flatter the curve, the larger the variance).

Ratcliff (1979) has described a method for combining data from individual subjects to produce group reaction time distributions. RTs for each subject are organized in ascending order, and quantiles are calculated. The quantiles are then averaged over subjects to give group quantiles. This is called Vincent-averaging or Vincentizing (Vincent, 1912). From the group quantiles a group RT-distribution can be constructed (see the example in Figure A.2, adapted from Ratcliff, 1979). Each cumulative relative frequency in this group distribution is associated with the mean of the RTs from all subjects for that particular quantile. The group distribution method averages over individual subjects' data in a way that retains shape information. For certain distributional forms (such as the normal) this results in a group distribution of the same functional form (necessary and sufficient conditions for this procedure to work are that the distributions belonging to the same family can all be standardized, e.g., differ from each other only with respect to scale and location, see Thomas & Ross, 1980). An advantage of this approach is that outliers can be kept till the last moment: They simply do not influence the general shape of the resulting distribution. They can be discarded afterwards.

In order to apply Ratcliff's method to empirical data, it is practical if all distributions involved are based on the same number of observations, so that a simple ordering of RTs suffices to determine which RTs go with which quantiles. Therefore, some method must be used to replace missing values, or to estimate the quantiles on the basis of the RTs that are available. In our experiments we have estimated the quantiles in two ways: by linear interpolation and by cubic interpolation (spline). No important deviations were found between the results of these two approaches.

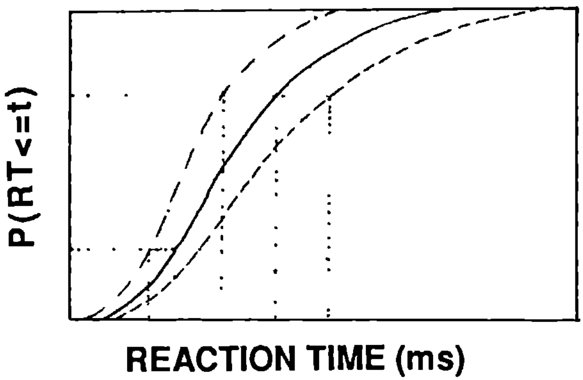


Figure A.2. An example of Vincent-averaging applied to cumulative distribution functions.

A1.2 Independent and dependent statistical facilitation

Figure A.3 (adapted from Ulrich and Giray, 1986, p. 250) summarizes the different possible situations arising from independent and dependent statistical facilitation. In this figure, $P(RT_A \leq t) + P(RT_B \leq t) - P(RT_A \leq t)P(RT_B \leq t)$ gives the cumulative distribution curve for two independent channels A and B; for dependent channels, $P(RT_A \leq t) + P(RT_B \leq t)$ represents Miller's curve for maximal statistical facilitation, while the maximum of $P(RT_A \leq t)$ and $P(RT_B \leq t)$ at any point gives the minimal statistical

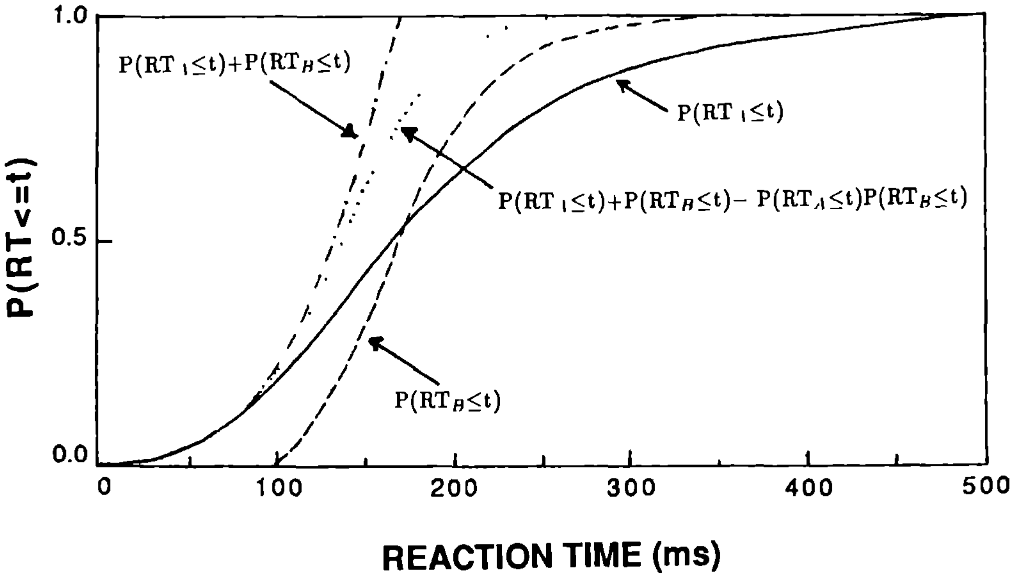


Figure A.3. Cumulative distribution functions for channels A, B and their minimum, assuming independence or dependence of the two channels.

facilitation possible.

To better understand this figure, let us reconsider Miller's (1982) formula for computing the cumulative minimum distribution $P(RT_{AB} \leq t)$ from two distributions $P(RT_A \leq t)$ and $P(RT_B \leq t)$ (cf. Chapter 4):

(1) $P(RT_{AB} \leq t) = P(RT_A \leq t) + P(RT_B \leq t) - P(RT_A \leq t \;\&\; RT_B \leq t)$

As we saw in Chapter 4, with independent channels the last term is equal to $P(RT_A \leq t)P(RT_B \leq t)$. However, the last term in the Equality can also be written in terms of conditional probabilities (Vorberg, personal communication):

$$\begin{aligned} P(RT_A \leq t \;\&\; RT_B \leq t) &= P(RT_A \leq t | RT_B \leq t)P(RT_B \leq t) \\ &= P(RT_B \leq t | RT_A \leq t)P(RT_A \leq t) \end{aligned}$$

The conditional probabilities are bounded by 0 and 1. The minimum of their product is 0, which turns the above Equation into Miller's sum-curve for maximal statistical facilitation:

$$P(RT_{AB} \leq t) \leq P(RT_A \leq t) + P(RT_B \leq t)$$

This situation will be approached if there is a high negative correlation between channels A and B (i.e., short RTs in A show up with long RTs in B). Furthermore, the product of the probabilities is maximally equal to 1, which leads to the following Inequality:

$$P(RT_{AB} \leq t) \geq \max (P(RT_A \leq t, P(RT_B \leq t))$$

This situation will be approached if there is a large positive correlation between channels A and B (i.e., long RTs in A show up with long RTs in B) (also see Ulrich & Giray, 1986; Colonius, 1987).

Instead of formula (1), Cohen (1984) and Massaro (1987) apply the following formula (2) to the situation of independent channels, which can be derived from (1) by differentiation (f indexes a density function, F a cumulative distribution function):

$$(2)\ f_{AB}(t) = f_A(t) + f_B(t) - (f_A(t)\, F_B(t) + F_A(t)\, f_B(t)), \text{ or}$$

$$f_{AB}(t) = f_A(t)\, (1-F_B(t)) + f_B(t)\, (1-F_A(t))$$

While formula (1) gives the relation between A, B and their minimum AB expressed as cumulative distributions, formula (2) uses the probability density function itself. Both formula (1) and (2) can be used to derive the predicted minimum RT for an independent race.

Instead of by using one of these formulas, one can also obtain the predicted RTs for the bimodal conditions by way of simulation. In this case the empirical distributions from the two single channel conditions are used as sources for random sampling. One RT is randomly sampled from distribution A and another from distribution B (and the value of the SOA is added to the appropriate channel); both are then compared and the fastest RT is selected. This is done a number of times, leading to a predicted distribution. The advantage of this method is that the modality from which a RT is sampled at each trial can be stored together with that RT: The resulting predicted RT can easily be split up in a certain proportion of A-reactions and of B-reactions, each having its own mean RT. This, of course, is one way to predict the visual and auditory RTs in the modality decision task under the race assumption. A disadvantage of this method is that, due to the random sampling, a certain variability is found in the resulting mean RTs when different sets of random trials are considered; outliers and amount of missing values in one or both distributions may influence the resulting distribution quite heavily. However, application of this method to the data from Chapters 4 and 5 did not lead to results that affect the interpretation given there in any way.

A1.3 Formulas used in the analysis of the modality decision task

In bimodal redundant conditions in the modality decision task, subjects reacted to the visual or to the auditory target (see Chapter 5). Assuming an independent race between the processes in the two channels, it is possible to derive the expected mean RT and proportion of reactions to each channel. In order to do so, we can use Bayes' formula:

$$P(F_j|E) = \frac{P(EF_j)}{P(E)} = \frac{P(E|F_j)P(F_j)}{\sum_i P(E|F_i)P(F_i)}$$

In our case:

$$P(A{=}t|A{<}B) = \frac{P(A{<}B|A{=}t)P(A{=}t)}{\sum_{t'} P(A{<}B|A{=}t')P(A{=}t')}$$

$$P(A{=}t|A{<}B) = \frac{P(B{>}t|A{=}t)P(A{=}t)}{\sum_{t'} P(B{>}t'|A{=}t')P(A{=}t')}$$

Given the independence of the channels A and B, $P(B{>}t|A{=}t) = P(B{>}t)$, leading to a simplification of the just presented equations. The mean RT to channel A, for those situations that the reactions to channel A were faster than to channel B, is found by multiplying all RTs with their relative frequency of occurrence and summing them:

$$\overline{\mathrm{RT}}_{A|A<B} = \sum_t t\mathrm{P}(\mathrm{A}=t|\mathrm{A}<\mathrm{B}).$$

To obtain the total proportion of reactions to channel A, given that A<B, the following derivation is possible (cf. Ross, 1988, p. 68):

$$\begin{aligned}\mathrm{P}(\mathrm{A}<\mathrm{B}) &= \sum_t \mathrm{P}(\mathrm{A}<\mathrm{B}\ \&\ \mathrm{A}=t) \\ &= \sum_t \mathrm{P}(\mathrm{A}<\mathrm{B}|\mathrm{A}=t)\mathrm{P}(\mathrm{A}=t).\end{aligned}$$

Taking into account that A and B are independent, this can be simplified to

$$\mathrm{P}(\mathrm{A}<\mathrm{B}) = \sum_t \mathrm{P}(\mathrm{B}>t)\mathrm{P}(\mathrm{A}=t).$$

A1.4 Parameter estimation and χ^2-test for goodness-of-fit

In order to test if a model gives a reasonable description of empirical data, some kind of a measure of fit between the two must be computed (cf. Wickens, 1982). One measure is the average absolute RT-difference between model and data (N indicates the number of datapoints involved):

$$\text{Av. Abs. Diff.} = \sum_i^{N} \frac{|\mathrm{RT}_{data_i} - \mathrm{RT}_{model_i}|}{n_i}$$

A second common stress-measure is the sum of the squared differences between the predicted and obtained RTs. This measure can be turned into an approximated χ^2 statistic by taking into account the sample variance s^2 in the obtained and predicted RTs as follows:

$$(3)\ \chi^2 = \sum_i^{N} \frac{(\mathrm{RT}_{data_i} - \mathrm{RT}_{model_i})^2}{\left(\frac{s^2{}_{data_i}}{n_i}\right) + \left(\frac{s^2{}_{model_i}}{m_i}\right)}$$

That this measure is approximately χ^2-distributed can be shown as follows (J. Havinga, personal communication). Let $\overline{Y}_i^D$ be the sample mean of data in cell i on the basis of n_i measurements Y_{ij}^D; let μ_i^D be the population mean of data in cell i; let $\sigma_i^{2\ D}$ be the population variance of data in cell

i. Analogously, let $\overline{Y}_i^M$ stand for the corresponding mean generated by the model on the basis of m_i simulations Y_{ij}^M, μ_i^M for the population mean and $\sigma_i^{2\ M}$ for the population variance of simulations in cell i. Then, following the Central Limit Theorem, for $n \rightarrow \infty$ (Mood, Graybill, & Boes, 1974, p. 234):

$$\overline{Y}_i^D = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}^D \stackrel{d}{=} N\left(\mu_i^D, \frac{\sigma_i^{2\ D}}{n_i} \right)$$

and

$$\overline{Y}_i^M = \frac{1}{m_i} \sum_{j=1}^{m_j} Y_{ij}^M \stackrel{d}{=} N\left(\mu_i^M, \frac{\sigma_i^{2\ M}}{m_i} \right)$$

If we now let $\overline{X}_i = \overline{Y}_i^D - \overline{Y}_i^M$ (keeping in mind that the goal of simulation is $\mu_i^D = \mu_i^M$), then

$$\mu_{X_i} = \mu_i^D - \mu_i^M = 0, \text{ and}$$

$$\sigma_{X_i}^2 = \frac{\sigma_i^{2\ D}}{n_i} + \frac{\sigma_i^{2\ M}}{m_i}$$

where $X_i \stackrel{d}{=} N(0, \sigma_{X_i}^2)$.

Furthermore, according to Mood et al., 1974, p. 241, Theorem 7):

$$(4)\ \chi_N^2 = \sum_{i=1}^{N} \frac{(X_i - \mu_{\bar{X}_i})^2}{\sigma_{X_i}^2}$$

By substitution of terms in (4) we obtain formula (3) given above, except that in (3) sample values are used as estimators for the population means and variances in the data and the model. This result is valid as long as for all i, j $Y_{i,j}^M$ and $Y_{i,j}^D$ are independent and identically distributed, and n_i and m_i are sufficiently large for all i.

When the model values are based on the use of p free parameters, $Y_{i,j}^M$ are no longer independent for all i,j. The number of degrees of freedom for the χ^2-distribution then becomes N-p, instead of N (where N indicates the number of data points involved).

Both small and large simulations were run. For the small simulations, sample size m_i was set equal to n_i (leading to a simplification of the formula); for the large simulations sample size m_i was set at 35,000.

For the computation of means and variances, data of all subjects were pooled. This has some consequences. As can be observed on the basis of Equation (3), considering subjects as replications on the one hand leads to an over-estimation of the population variance (due to subject effects), and thus results in a denominator that is too large. In fact, the variance in the pooled data was much larger than in the model. On the other hand, pooling leads to a much larger n_i, which results in a smaller denominator. The first consideration alone would lead to a χ^2 that is smaller than it should be and to too few rejections of the model; the second leads to too many rejections. A better, but more time consuming, approach would be to fit the model to each individual subject, compute the χ^2 for each subject, and sum the χ^2s over subjects, while adapting the degrees of freedom.

Appendix 2: Tables

This appendix contains Tables with the mean RTs and percentages of reactions for all experimental conditions that were presented in terms of Figures in Chapters 3, 4, 5, and 7.

Table associated with Figure 3.1.

	SOA	−190	−70	−30	+30	+150
CONDITION						
consonant-congruent		504	529	541	549	571
consonant-incongruent		521	541	549	564	582

Mean reaction times (in ms) for consonant-congruent and -incongruent conditions as a function of SOA.

Table associated with Figure 3.2.

CONDITION	SOA −190	−70	−30	+30	+150	
consonant-congruent	509	541	540	552	561	
consonant-incongruent	525	557	553	571	571	
single-channel						558

Mean reaction times (in ms) for the single-channel condition, and for the consonant-congruent and consonant-incongruent conditions as a function of SOA.

Table associated with Figure 3.3.

CONDITION	SOA −250	−100	0	+100	+250	
consonant-congruent	415	428	433	426	432	
consonant-incongruent	419	423	436	426	443	
star	421	424	435	429	440	
single-channel						444

Mean reaction times (in ms) for the single-channel condition, and for the consonant-congruent, -incongruent, and star-conditions as a function of SOA.

Table associated with Figure 3.4.

SOA	−250	−100	0	+100	+250	
CONDITION						
vowel-congruent	384	385	411	424	433	
vowel-incongruent	433	432	449	424	431	
star	409	415	428	424	433	
single-channel						437

Mean reaction times (in ms) for the single-channel condition, and for the vowel-congruent, -incongruent and star-conditions as a function of SOA.

Table associated with Figure 3.5.

SOA	−250	−100	0	+100	+250	single-channel
CONDITION						
CV-syllable	443	439	456	446	453	460
VC-syllable	394	401	418	415	428	429
V-syllable	389	392	413	410	417	423

Mean reaction times (in ms) for CV-, V- and VC-syllable types in the single-channel condition, and in the combined vowel-congruent, -incongruent and star-conditions as a function of SOA.

Table associated with Figure 4.4.

CONDITION	s.c.v.	SOA1=−100	SOA2=0	SOA3=100	s.c.a.
Aa	406	375	334	370	396
Au	406	408	379	403	443
Ua	422	410	364	409	396
Uu	422	402	369	399	443

Mean RTs (in ms) in the redundant and single-channel conditions when the visual stimulus preceded the auditory stimulus (SOA1=−100 ms), accompanied it (SOA2=0 ms), or followed it (SOA3=100 ms). RTs are measured from the first target stimulus. S.c.v. stands for "single-channel visual", s.c.a. for "single-channel auditory".

Table associated with Figures 4.6 and 4.7.

CONDITION	s.c.v.	SOA1=−100	SOA2=0	SOA3=100	s.c.a.
Ai	406	407	419	417	–
Ui	422	428	426	437	–
An	406	386	376	380	–
Un	422	402	390	395	–
Ia	–	407	404	411	396
Iu	–	440	448	421	443
*a	–	386	401	409	396
*u	–	413	441	432	443

Mean RTs (in ms) in the non-redundant and single-channel conditions under three SOAs. S.c.v. stands for "single-channel visual", s.c.a. for "single-channel auditory"; n stands for "NOISE".

Table 1 associated with Figure 5.1.

OBTAINED AND PREDICTED PERCENTAGE OF REACTIONS TO VISUAL TARGET

		v first						a first
	SOA	−100	−40	−20	0	20	40	100
CONDITION								
Aa	OBTAINED	68.4	64.2	59.4	58.1	57.2	53.6	37.9
	PREDICTED	85.6	71.0	65.4	59.8	53.6	47.6	30.6
	DIFFERENCE	17.2	6.8	6.0	1.7	−3.6	−6.0	−7.3
Au	OBTAINED	82.2	73.4	69.9	66.5	58.1	58.0	37.6
	PREDICTED	90.9	80.4	75.0	69.0	62.5	55.6	35.9
	DIFFERENCE	8.7	7.0	5.1	2.5	4.4	−2.4	−1.7
Ua	OBTAINED	70.6	58.8	56.6	49.0	47.4	40.1	32.7
	PREDICTED	80.3	65.3	59.6	54.1	48.2	42.8	27.5
	DIFFERENCE	9.7	6.5	3.0	5.1	0.8	2.7	−5.2
Uu	OBTAINED	75.5	70.7	66.8	63.0	60.3	50.5	41.1
	PREDICTED	87.5	75.1	69.2	63.2	56.4	49.4	31.4
	DIFFERENCE	12.0	4.4	2.4	0.2	−3.9	−1.1	−9.7

Obtained and predicted mean percentage of reactions to the visual target, and their difference, in the redundant conditions under all SOAs.

Table 2 associated with Figure 5.1.

OBTAINED AND PREDICTED PERCENTAGE OF REACTIONS TO AUDITORY TARGET

		v first						a first
	SOA	−100	−40	−20	0	20	40	100
CONDITION								
Aa	OBTAINED	29.6	33.6	38.0	40.7	41.4	43.8	58.2
	PREDICTED	14.6	29.2	34.7	40.6	46.8	52.6	69.5
	DIFFERENCE	−15.0	−4.4	−3.3	−0.1	5.4	8.8	11.3
Ua	OBTAINED	26.6	39.1	41.8	48.5	49.4	56.4	62.9
	PREDICTED	19.9	35.1	40.6	46.3	52.0	57.6	73.0
	DIFFERENCE	−6.7	−4.0	−1.2	−2.2	2.6	1.2	10.1
Au	OBTAINED	15.5	25.0	28.1	31.7	39.5	38.4	57.6
	PREDICTED	9.3	19.9	25.4	31.3	38.0	44.9	64.6
	DIFFERENCE	−6.2	−5.1	−2.7	−0.4	−1.5	6.5	7.0
Uu	OBTAINED	22.2	27.1	30.8	35.5	38.1	46.1	55.1
	PREDICTED	12.7	25.1	31.0	37.4	44.4	50.9	69.0
	DIFFERENCE	−9.5	−2.0	0.2	1.9	6.3	4.8	13.9

Obtained and predicted mean percentage of reactions to the auditory target, and their difference, in the redundant conditions under all SOAs.

Table 1 associated with Figure 5.2.

REACTIONS TO VISUAL TARGET

			v first						a first
	SOA	s.c.v.	−100	−40	−20	0	20	40	100
CONDITION									
Aa	OBTAINED	397	370	351	347	344	362	375	430
	PREDICTED		382	376	373	371	388	404	456
	DIFFERENCE		12	25	16	27	26	29	26
Au	OBTAINED	397	389	371	368	363	386	395	462
	PREDICTED		384	378	375	372	389	406	456
	DIFFERENCE		-5	7	7	9	3	11	-6
Ua	OBTAINED	413	386	380	380	367	377	388	456
	PREDICTED		394	386	383	379	396	412	462
	DIFFERENCE		8	6	3	12	19	24	6
Uu	OBTAINED	413	393	381	359	355	388	399	450
	PREDICTED		397	389	386	382	399	414	463
	DIFFERENCE		4	8	27	27	11	15	13

Obtained and predicted mean RTs to visual targets, and their differences (in ms) in the redundant conditions under all SOAs. RTs were measured from the onset of the first presented target stimulus. Single-channel RTs (s.c.) are also indicated. The means are based on different numbers of correct reactions for each subject.

Table 2 associated with Figure 5.2.

REACTIONS TO AUDITORY TARGET

		v first					a first		
	SOA	−100	−40	−20	0	20	40	100	s.c.a.
CONDITION									
Aa	OBTAINED	363	323	309	296	300	326	353	434
	PREDICTED	424	383	370	357	365	372	390	
	DIFFERENCE	61	60	61	61	65	46	27	
Ua	OBTAINED	386	345	323	313	328	345	373	434
	PREDICTED	438	394	380	367	373	379	394	
	DIFFERENCE	52	49	57	46	45	34	21	
Au	OBTAINED	411	358	343	341	338	352	397	456
	PREDICTED	445	404	391	379	386	393	411	
	DIFFERENCE	33	46	48	38	48	41	14	
Uu	OBTAINED	412	358	353	345	360	354	389	456
	PREDICTED	459	416	402	389	395	401	416	
	DIFFERENCE	47	58	49	44	35	47	27	

Obtained and predicted mean RTs to auditory targets, and their differences (in ms) in the redundant conditions under all SOAs. RTs were measured from the onset of the first presented target stimulus. Single-channel RTs (s.c.) are also indicated. The means are based on different numbers of correct reactions for each subject.

Table associated with Figures 5.3 and 5.4.

		v first					a first		
SOA:	s.c.v.	−100	−40	−20	0	20	40	100	s.c.a.
CONDITION									
Ai	397	406	414	399	393	397	408	410	
Ui	413	431	420	409	413	420	418	406	
Ia		355	378	399	395	405	405	414	434
Iu		393	419	422	435	429	435	436	456

Mean RTs (in ms) for eight subjects in the four bimodal non-redundant conditions under seven SOAs. Single-channel RTs (s.c.) are also indicated.

Table associated with Figures 7.2 and 7.3.

INCONGRUENT CONDITIONS

EXP	COND	SOA	VIS	AUD	RED	MODEL
4	Ua	−100	393	386	378	378
4	&a	−100	372	376	362	362
5	Ua	−100	376	348	360	360
5	*a	−100	368	351	351	355
5	Ua	0	373	351	317	318
5	*a	0	358	360	310	317
6	Au	−100	406	443	408	399
6	Ua	−100	422	396	410	403
6	Au	0	406	443	379	374
6	Ua	0	422	396	364	360
6	Au	100	406	443	403	417
6	Ua	100	422	396	409	385

CONGRUENT CONDITIONS

EXP	COND	SOA	VIS	AUD	RED	MODEL
4	Aa	−100	334	349	315	321
5	Aa	−100	345	345	328	328
5	Aa	0	340	345	290	285
6	Aa	−100	406	396	375	380
6	Uu	−100	422	443	402	399
6	Aa	0	406	396	334	334
6	Uu	0	422	443	369	360
6	Aa	100	406	396	370	374
6	Uu	100	422	443	399	407

Parameter values: n_v=25, n_a=22, α=.76

This Table shows the visual (VIS) and auditory (AUD) single-channel mean RTs that served as input to the model as well as the empirically obtained bimodal redundant mean RTs (RED). Further shown are the predictions for these redundant RTs given by a simulation with 35,000 samples per data point (MODEL).

Table associated with Figures 7.4, 7.5, and 7.6.

SOA	VIS	AUD	RED		MODEL		
			Aa	Ee			
Congruent							
−250	(423)	423	364	358	361		
−100			351	372	372		
0			388	401	382		
100			405	419	400		
250			407	430	418		
			Ea	Ae	MODEL		
Incongruent							
−250	(423)	423	416	428	411		
−100			404	429	422		
0			424	439	424		
100			397	415	420		
250			412	415	421		
			*a	*e	Pa	Pe	MODEL
Neutral							
−250	(423)	423	378	387	394	379	382
−100			399	397	413	407	395
0			403	425	402	415	405
100			408	417	408	435	412
250			424	413	385	438	419

Parameter values: n_v=25, n_a=22, α=.76, δ=.91, ι=1.21

Obtained (RED) and predicted (MODEL) mean RTs for the vowel-congruent, -incongruent and bimodal neutral conditions in auditory vowel-detection Experiment 3, together with parameter values and statistics (VIS= visual single-channel RT, AUD= auditory single-channel RT, RED= redundant RT).

Notes

Note to Chapter 1

1. For convenience, I use the term "vowel" to refer to both visual and auditory representations (e.g., A and /a:/). The term "consonant" is used in an abstract a-modal way as well (e.g., P and /p/).

Notes to Chapter 2

1. According to Webster's New Collegiate Dictionary, "phonetic" is a term used in relation to speech sounds and spoken language. Indeed, Perfetti, Bell, and Delaney (1988) use the term "phonetic" (in contrast with "graphic") as associated with the speech form of a written word. Posner and Hanson (1981), however, use the term "in a loose sense to refer to the segmental structure of words". They state that it does not imply fine acoustic and phonetic distinctions such as the differences in /t/ in /tip/, /step/, and /pit/. Other researchers (e.g., Campbell, 1987, p. 146) sometimes use the term "phonetic" to refer to an abstract amodal code.

2. Posner and Hanson (1981) actually seem to agree with this conclusion, since they specifically state (p. 208): "What is lacking, however, is convincing evidence that the phonetic code used for processing visual words is the same as the phonetic code used for processing auditory words".

3. An interesting aspect of the pattern of results Boies obtained is that the memory-load conditions were *faster* than the no memory-load conditions at an ISI of 0 ms. This could be interpreted as an effect of arousal caused by the presence of a stimulus in the second channel. Furthermore, most of the studies mentioned here show a relevant contribution of the acoustic code only after relatively long ISIs. In this respect these studies seem more relevant to how information is stored in a short-term memory buffer than to fast cross-modal activation, the issue which is addressed here.

4. Greenwald's results point to a difference between conditions with and without a neutral stimulus (tap). They indicate that irrelevant material is processed

even though subjects were instructed to ignore auditory input as best as they could (cf. Chapter 3 of this thesis). Also, the auditory letter condition showed differences between letter and different digit conditions only in terms of error rate, which could be interpreted as the absence of inhibition at a representation level (cf. Chapter 4, Experiment 6).

Notes to Chapter 3

* This chapter has been published in a slightly different form in Language and Speech, 1989, 32, 89-108, under the authorship of Dijkstra, Schreuder and Frauenfelder. This explains the use of the pronoun "we" in this chapter as a deviation from the pronoun "I" in other chapters.

1. The term "phonetic" is often used to indicate physical characteristics of the speech signal. We therefore prefer to use the term "phonological" to refer to abstract representations in the auditory modality in general, and the term "phonemic" when phoneme representations are involved. In the same fashion we use the terms "orthographic" and "graphemic" when talking about abstract representations in the visual domain.

2. Another mechanism that does not assume that consonants and vowels are processed as integral units can be found in the TRACE I model of auditory word recognition (Elman & McClelland, 1986). In this model units representing stops are allowed to modify connections between units for features and vowel units that follow. Thus, a consonant like [p] would change the connection between a vowel like [a] and certain features of it (e.g., formant values), facilitating the recognition of the vowel by compensating for the contextual effects observed in the signal.

3. In a pilot run we had subjects *identify* the letter that appeared in the preceding trial. Since they kept confusing the presented letter with the auditory target syllable, in the experiment we only asked whether any visual stimulus had been presented at all. Still, the confusion observed already seems to indicate that some kind of automatic cross-modal activation was going on.

4. The vowel /e:/ was chosen for two main reasons. First, we hoped to avoid deviating results like that for the syllable /ko/ in Experiment 1. Second, Experiments 2 and 3 also involved a manipulation of letter name vs. abstract letter representation. E.g., in Dutch the letter name for P is /pe:/ and for K is /ka:/. By using target syllables like /pa:/, /pe:/ and /ka:/, /ke:/ the effect of letter name vs. abstract consonant representation on the auditory syllable could be distinguished. Since this manipulation did not lead to consistent results over syllables and experiments, the effects of this factor were not further documented.

5. The largest RT-differences between consonant-congruent and -incongruent conditions were found for /pa:/, the smallest for /pe:/. The size of the congruence effect for each SOA varied somewhat over syllables, but there was no particular syllable that deviated consistently from the others over the whole SOA-range.

6. While the syllables in Experiment 1 were all clearly nonwords, some of those in Experiment 2 could in fact be treated as words by the subjects (e.g., /pa:/ and /te:/). Since many CV-syllables have lexical status, this is hard to avoid. Because the syllables containing a /k/-consonant were all either nonwords or words of a negligible frequency, this consonant was chosen for the CV- and VC-syllables in Experiment 3. Furthermore, no clear RT-differences were found in Experiments 2 and 3 between syllables that could be considered words or not.

Notes to Chapter 4

1. I use the notation /u:/ for the phoneme that is pronounced as in French *lune* or in German *über*, as a concession to the visual notation U. In the International Phonetic Alphabet (IPA) it would be written as /y/. The notation used for other phonemes and syllables in this thesis approximates that of the IPA.

2. Recently, some problems with the interpretation of data obtained with the bimodal detection task have been noted, particularly in terms of fast guessing (Eriksen, 1988) and response preference strategies (Mullin, Egeth, & Mordkoff, 1988). However, it seems reasonable to assume that these general strategies should influence my results similarly for congruent and incongruent conditions and thus be of little consequence for my interpretation.

3. Because of the varying characteristics of the single-channel distributions, it is strictly spoken impossible to equate the congruent and incongruent redundant conditions in all respects (e.g., in amount of overlap of distributions). However, there is no a priori reason to assume that congruent conditions would consistently (over SOAs and experiments) benefit from such a factor and incongruent conditions would not. Indeed, since the temporal relation between distributions changes over SOA, if the congruent conditions would benefit from more overlap under one SOA, they should necessarily be at a disadvantage under others. Also, averaging over subjects should diminish this problem. As the chapter will show, the analysis method used leads to consistent results over the experiments presented, though these differed in terms of their design (mixed or blocked), instructions and test conditions.

4. Though not reported, I also tested differences in the amount of coactivation between congruent and incongruent conditions by means of an estimation of the surface violation present in both cases (following Miller, 1986, p. 336). The results of these tests always led to the same conclusions as those in the text, which were based on the assumption of independent channels. Similar

conclusions are also reached when congruent and incongruent conditions are compared in terms of their RT-difference with respect to the fastest single-channel components (taking SOA into account).

5. To keep the notation consistent with that of Chapter 3, a negative SOA indicates a condition where a visual stimulus precedes an auditory stimulus, while a positive SOA indicates a condition in which an auditory stimulus leads. Reaction time, however, is always measured from the onset of the first presented target (either visual or auditory). E.g., an SOA of -100 ms for condition Ia stands for a condition in which visual stimulus I precedes auditory stimulus /a./ by 100 ms. If I is not a target, but /a:/ is, RT is measured from the onset of /a./.

6. The reader will have noticed that the hypothesis of a "race to recognition" between the visual and auditory targets is also in agreement with the relatively fast RTs at SOA2=0 ms compared to those at SOA1=-100 ms.

Notes to Chapter 5

1. The relatively large deviations between the obtained and predicted means for auditory reactions are to some extent misleading. To understand this, it must be noted that the obtained means used in Figure 5.2 are based on unequal numbers of RTs per subject, depending on how often they reacted to the visual and to the auditory modality. However, the predicted means are based on equal numbers, namely the maximum number of reactions that could be given (90). If the obtained means are subtracted from the predicted means for each subject *individually* (as described in the text), the average deviation over subjects now is only 24 ms facilitation for auditory and 18 ms facilitation for visual reactions.

2. Further experiments with the modality decision task could try to diminish the noise by asking the subject after each trial to indicate whether the response given was really intended. In my experiment this solution was not chosen because it is has some drawbacks. It increases the length of the sessions, it breaks the rhythm of responding by the subject, and it could be considered as introducing a second task (that of temporal order judgment). For some types of questions, the problem of wrong responses may be avoided by using a naming task instead of a two-choice response task (cf. Hell, 1987). However, in my situation I would not have been able to analyze the congruent conditions with respect to modality.

3. A similar waiting strategy could have occurred in the bimodal detection experiments of Chapter 4. However, subjects participating in those experiments never reported (an experience of) waiting for the second stimulus, and the obtained results show no sign of it. Also note that such a strategy would not

consistently favor the congruent conditions above the incongruent conditions (advantages at one SOA turn into disadvantages at other SOAs).

4. Theories about the allocation of attention in bimodal tasks would probably give a more sophisticated account. For example, Duncan (1980) assumes that processing in the visual and auditory channels up to a certain point proceeds in parallel without attention decrement. After this point, targets compete for admission to a limited-capacity system for further processing, while non-targets are filtered out.

References

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 802-814.

Appelman, I.B., & Mayzner, M.S. (1981). The letter-frequency effect and the generality of familiarity effects on perception. *Perception and Psychophysics, 30*, 436-446.

Black, M., Coltheart, M., & Byng, S. (1987). Forms of coding in sentence comprehension during reading. In M. Coltheart (Ed.), *Attention and Performance XII.* Hillsdale (N.J.): Lawrence Erlbaum Associates, 655-672.

Blosfelds, M. (1981). *Visual and auditory word recognition.* Unpublished honours thesis, Monash University.

Boies, S.J. (1971). *Memory codes in a speeded classification task.* Unpublished doctoral dissertation, University of Oregon.

Boles, D.B. (1981). Variability in letter matching asymmetry. *Perception and Psychophysics, 29*, 285-288.

Boles, D.B., & Eveland (1983). Visual and phonetic codes and the process of generation in letter matchings. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 657-674.

Bower, G.H. (1975) Cognitive psychology: An introduction. In: W.K. Estes (Ed.), *Handbook of learning and cognitive processes I.* Hillsdale (N.J.): Lawrence Erlbaum Associates, 25-80.

Bradley, D.C., & Forster, K.I. (1987). A reader's view of listening. *Cognition, 25*, 103-134.

Campbell, R. (1987). Common processes in immediate memory: Precategorical acoustic storage and some of its problems. In A. Allport, D.G. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Relationships between listening, speaking, reading and writing.* New York: Academic Press, 131-150.

Carlton, L.G. (1981). Processing visual feedback information for movement control. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1019-1030.

Carrasco, M., Kinchla, R.A., & Figueroa, J.G. (1988). Visual letter-matching and the time course of visual and acoustic codes. *Acta Psychologica, 69*, 1-17.

Chase, W.G., & Calfee, R.G. (1969). Modality and similarity effects in short-term recognition memory. *Journal of Experimental Psychology, 81*, 510-514.

Cohen, G. (1968). A comparison of semantic, acoustic, and visual criteria for matching of word pairs. *Perception and Psychophysics, 3*, 203-204.

Cohen, G. (1969). Some evidence for parallel comparisons in a letter recognition task. *Quarterly Journal of Experimental Psychology, 21*, 272-279.

Cohen, M.M. (1984). Processing of visual and auditory information in speech perception. Unpublished doctoral dissertation. Santa Cruz: University of California.

Colavita, F.B. (1974). Human sensory dominance. *Perception and Psychophysics, 16*, 409-412.

Colonius, H. (1987). Modeling dependent processing in reaction time analysis. In Roskam, E.E. & Suck, R. (Eds.), *Progress in Mathematical Psychology.* Elsevier Science Publishers BV., vol. I.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of Information Processing*, New York: Academic Press, 151-216.

Coltheart, M. (1980). Reading, phonological recoding and deep dyslexia. In M. Coltheart, K.E. Patterson & J.C. Marshall (Eds.), *Deep Dyslexia*, London: Routledge and Kegan Paul, 197-226.

Coltheart, M., Davelaar, E., Jonasson, J.T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI*, Hillsdale (N.J.): Lawrence Erlbaum Associates, 535-555.

Dainoff, M.J. (1970). Time-course of visual and auditory encoding. *Journal of Experimental Psychology, 86*, 214-224.

Dainoff, M.J., & Haber, R.N. (1967). How much help do repeated presentations give to recognition processes? *Perception and Psychophysics, 2*, 131-136.

Dainoff, M.J., & Haber, R.N. (1970). Effect of acoustic confusability on levels of information processing. *Canadian Journal of Psychology, 24*, 98-108.

De Boor, C. (1978). A practical guide to splines. New York: Springer Verlag.

Donders, F.C. (1868-69/1969). Over de snelheid van psychische processen. (On the speed of mental processes). *Onderzoekingen gedaan in het physiologisch Laboratorium der Utrechtsche Hoogeschool*, Tweede reeks, 11, 92-130. Translated by W.G. Koster, in W.G. Koster (Ed.), Attention and Performance II, *Acta Psychologica, 30*, 412-431.

Donnenwerth-Nolan, S., Tanenhaus, M.K., & Seidenberg, M.S. (1981). Multiple code activation in word recognition: Evidence from rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 170-180.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review, 87*, 272-300.

Egeth, H.E., & Sagar, L.C. (1977). On the locus of visual dominance. *Perception and Psychophysics, 22*, 77-86.

Ehri, L.C. (1985). Effects of printed language acquisition on speech. In D.R. Olson, N. Torrance, & A. Hildegard (Eds.), *Literacy, language and learning: The nature and consequences of reading and writing.* Cambridge: Cambridge University Press.

Elman, J.E., & McClelland, J.L. (1986). Exploiting lawful variability in the speech wave. In J.S. Perkell & D H. Klatt (Eds.), *Invariance and Variability in Speech Processes*, Hillsdale (N.J.): Lawrence Erlbaum Associates, 360-385.

Eriksen, C.W. (1988). A source of error in attempts to distinguish coactivation from separate activation in the perception of redundant targets. *Perception and Psychophysics, 44*, 191-193

Eriksen, B.A., & Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics, 16*, 143-149.

Evett, L.J., & Humphreys, G.W. (1981). The use of abstract graphemic information in lexical access. *Quarterly Journal of Experimental Psychology, 33A*, 325-350.

Flores d'Arcais, G.B., Schreuder, R., & Glazenborg, G. (1985). Semantic activation during recognition of referential words. *Psychological Research, 37*, 39-49.

Fodor, J.A. (1975). *The language of thought.* New York: Thomas Y. Crowell.

Fodor, J.A. (1981). *Representations.* Cambridge (Mass.). MIT Press.

Fodor, J.A. (1983). *The modularity of mind.* Cambridge (Mass.): MIT Press.

Forster, K.I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 680-698.

Frauenfelder, U.H. (1990). Structure and process in the human mental lexicon. In H. Haken (Ed.), *The synergetics of cognition.* Hamburg: Springer Verlag.

Frauenfelder, U.H., Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experiment Psychology: Human Perception and Performance*, in press.

Frost, R., & Katz, L. (1989). Orthographic depth and the interaction of visual and auditory processing in word recognition. *Memory & Cognition, 17*, 302-310.

Frost, R., Repp, B.H., & Katz, L. (1988). Can speech perception be influenced by simultaneous presentation of print? *Journal of Memory and Language, 27*, 741-755.

Gaillard, A.W.K., & Perdok, J. (1980). Slow brain potentials in the CNV-paradigm. *Acta Psychologica, 44*, 147-163.

Glaser, M.O., & Glaser, W.R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 875-894.

Glaser, W.R., & Düngelhoff, F-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 640-654

Glucksberg, S., Fisher, D.G., & Monty, R.A (1967). Brief visual memory as a function of visual and auditory confusability Paper presented at the meeting of the American Psychological Association, Washington, D.C., September

Goldinger, S.D., Luce, P.A., & Pisoni, D.B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language, 28*, 501-518.

Gordon, P.C., & Meyer, D.E. (1984). Perceptual-motor processing of phonetic features in speech. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 153-178.

Gottlieb, G.L. & Agarwal, G.C. (1973). Response to sudden torques about ankle in man: Myotatic reflex. *Journal of Neurophysiology, 42*, 91-106.

Grainger, J. & Dijkstra, A. (in preparation). How monolingual can a bilingual be: Interlexical interference in bilingual word recognition.

Grainger, J., O'Regan, J.K., Jacobs, A.M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception & Psychophysics, 45*, 189-195.

Greenwald, A.G. (1970a). A double stimulation test of ideomotor theory with implications for selective attention. *Journal of Experimental Psychology, 84*, 392-398.

Greenwald, A.G. (1970b). Selective attention as a function of signal rate. *Journal of Experimental Psychology, 86*, 48-52.

Greenwald, A.G. (1970c). A choice reaction time test of ideomotor theory. *Journal of Experimental Psychology, 86*, 20-25.

Hanson, V.L. (1981). Processing of written and spoken words: Evidence for common coding. *Memory & Cognition, 9*, 93-100.

Hardzinski, M., & Pachella, R.G. (1980). The manipulation of stimulus quality and the definition of stimulus encoding operations in memory scanning experiments. *Perception & Psychophysics, 27*, 232-240.

Healy, A.F., & Cutting, J.E. (1976). Units of speech perception: Phoneme and syllable. *Journal of Verbal Learning and Verbal Behavior, 15*, 73-83.

Hell, W. (1987). *Aufmerksamkeit und Interferenz. Untersuchungen zur kontextabhängigen Verarbeitung von Ziffern.* Weinheim: Deutscher Studienverlag.

Humphreys, G.W., & Evett, L.J. (1985). Are there independent lexical and nonlexical routes in word processing? An evaluation of the dual-route theory of reading. *The Behavioral and Brain Sciences, 8*, 689-740.

Jackson, A., & Morton, J. (1984). Facilitation of auditory word recognition. *Memory & Cognition, 12*, 568-574.

Jakimik, J., Cole, R.A., & Rudnicky, A.I. (1985). Sound and spelling in spoken word recognition. *Journal of Memory and Language, 24*, 165-178.

James, W. (1890). *The principles of psychology.* New York: Dover Publications, vol. I, 488.

Johnston, W.A. & Wilson, J. (1980). Perceptual processing of non-targets in an attention task. *Memory & Cognition, 8*, 372-377.

Kahneman, D., & Chajczyk, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 497-509.

Kaplan, G.A., Yonas, A., & Shurcliff, A. (1966). Visual and acoustic confusability in a visual search task. *Perception and Psychophysics, 1*, 172-174.

Keuss, P.J.G. (1987). Intersensorische facilitatie. In Beishuizen, J.J., de Leeuw, L., & de Swart, J.H. (Eds.), *Cognitieve psychologie in theorie en praktijk.* Lisse: Swets & Zeitlinger, 39-59.

Kinchla, R. (1974). Detecting target elements in multielement arrays: A confusability model. *Perception and Psychophysics, 15*, 149-158.

Kirsner, K., Milech, D. & Standen, P. (1983). Common and modality-specific processes in the mental lexicon. *Memory & Cognition, 11*, 621-630.

Kosslyn, S.M. (1981). The medium and the message in mental imagery: A theory. *Psychological Review, 88*, 46-66.

LaHeij, W. (1988). Components of Stroop-like interference in picture naming. *Memory & Cognition, 16*, 400-410.

Levelt, W.J.M., Schriefers, H., Vorberg, D., Meyer, A., Pechmann, T., & Havinga, J. (submitted). The time course of lexical access in object naming.

Liberman, A.M. (1988). Reading is hard just because listening is easy. *Status Report on Speech Research, SR-95/96.* New Haven (Conn.): Haskins Laboratories.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

Luce, R.D. (1986). *Response times: Their role in inferring elementary mental organization.* Oxford: Oxford University Press.

Lukatela, G., Turvey, M.T., Feldman, L.B., Carello, C., & Katz, L. (1989). Alphabet priming in bi-alphabetical word perception. *Journal of Memory and Language, 28*, 237-254.

Maki, D.P. & Thompson, M. (1973). *Mathematical models and applications: With emphasis on the social, life and management sciences.* Englewood Cliffs (N.J.): Prentice-Hall.

Marcel, A.J. (1980). Surface dyslexia and beginning reading: A revised hypothesis of the pronunciation of print and its impairments. In M. Coltheart, K.E. Patterson, & J.C. Marshall (Eds.), *Deep Dyslexia*, London: Routledge and Kegan Paul, 227-258.

Marshall, J.C. (1984). Multiple perspectives on modularity. *Cognition, 17*, 209-242.

Marslen-Wilson, W.D. (1990). Activation, competition, and frequency in lexical access: Psycholinguistic and computational perspectives. In G. Altmann (Ed.), *Cognitive models of speech processing.* Cambridge (Mass.): MIT Press.

Martin, M. (1980). Attention to words in different modalities: Four-channel presentation with physical and semantic selection. *Acta Psychologica, 44*, 99-115.

Massaro, D.W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale (N.J.): Lawrence Erlbaum Associates.

McClelland, J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375-407.

McGill, W.J. (1963). Stochastic latency mechanisms. In: R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology vol. I*, 309-360.

Meyer, D.E., & Gutschera, K.D. (1975, November). Orthographic versus phonemic processing of printed words. Paper presented at the meeting of The Psychonomic Society, Denver.

Miller, J. (1978). Multidimensional same-different judgements: Evidence against independent comparisons of dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 411-422.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14*, 247-279.

Miller, J. (1985). Discrete and continuous models of divided attention. In M.I. Posner & O.S.M. Marin (Eds.), *Attention and Performance XI*, Hillsdale (N.J.): Lawrence Erlbaum Associates, 513-528.

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception and Psychophysics, 40 (5)*, 331-343.

Monsell, S. (1985). Repetition and the lexicon. In Ellis, A. (ed.) *Progress in the psychology of language.* London: Lawrence Erlbaum Associates, vol II.

Monsell, S., & Banich, M.T. (1982). Lexical priming: Repetition effects across input and output modalities. Paper presented at the 23rd annual meeting of the Psychonomic Society in Minneapolis.

Mood, A.M., Graybill, F.A., & Boes, D.C. (1974/1988). *Introduction to the theory of statistics.* Tokyo: McGrawhill.

Mullin, P.A., Egeth, H.E. & Mordkoff, J.T. (1988). Redundant-target detection and processing capacity: The problem of positional preferences. *Perception and Psychophysics, 43*, 607-610.

Mynatt, B.T. (1977). Reaction times in a bisensory task: Implications for attention and speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 316-324.

Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*, 226-254.

Nickerson, R.S. (1970). The effect of preceding and following auditory stimuli on response times to visual stimuli. *Acta Psychologica, 30*, 5-20.

Nickerson, R.S. (1973). Intersensory facilitation of reaction time: Energy summation or preparation enhancement? *Psychological Review, 80*, 489-509.

Nickerson, R.S. (1981). Speech understanding and reading: Some differences and similarities. In O.J.L. Tzeng & H. Singer (Eds.), *Perception of Print.* Hillsdale (N.J.): Lawrence Erlbaum Associates, 257-289.

Norris, D. (1986). Word recognition: Context effects without priming. *Cognition, 22*, 93-136.

Patterson, K., & Coltheart, V. (1987). Phonological processes in reading. In M. Coltheart (Ed.), *Attention and Performance XII.* Hillsdale (N.J.): Lawrence Erlbaum Associates, 421-447.

Perfetti, C.A., Bell, L.C., & Delaney, S.M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27*, 59-70.

Phaf, R.H. (1986). A connectionist model for attention: Restricting parallel processing through modularity. Unpublished thesis, University of Leiden.

Phaf, R.H., van der Heijden, A.H.C., & Hudson, P.T.W. (1990). SLAM: a connectionist model for attention in visual selection tasks. *Cognitive Psychology*, in press.

Posner, M.I. (1969). Abstraction and the process of recognition. In G. Bower & J.T. Spence (Eds.), *Psychology of learning and motivation III.* New York: Academic Press.

Posner, M.I. (1978). *Chronometric Explorations of Mind.* Hillsdale (N.J.): Lawrence Erlbaum Associates.

Posner, M.I., & Mitchell, R.F. (1967). Chronometric analysis of classification. *Psychological Review, 74*, 392-409.

Posner, M.I., Boies, S.J., Eichelman, W.H. & Taylor, R.L. (1969). Retention of visual and name codes of single letters. *Journal of Experimental Psychology, Monograph, 79*, 1-16.

Posner, M.I., & Snyder, C.R.R. (1975a). Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: the Loyola symposium*, Hillsdale (N.J.): Lawrence Erlbaum Associates, 55-85.

Posner, M.I., & Snyder, C.R.R. (1975b). Facilitation and inhibition in the processing of signals. In P.M.A. Rabbitt & S. Dornic (Eds.), *Attention and Performance V.* New York: Academic Press.

Posner, M.I., Nissen, M.J., & Klein, R.M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review, 83*, 2, 157-171.

Posner, M.I., & Hanson, V.L. (1981). Analyzing spoken and written language. In Cole, R.A. (Ed.), *Perception and production of fluent speech.* Hillsdale (N.J.): Lawrence Erlbaum Associates, 201-211.

Poulton, E.C. (1973). Unwanted range effects from using within subjects experimental designs. *Psychological Bulletin, 80*, 113-121.

Proctor, R.W. (1981). A unified theory for matching task phenomena. *Psychological Review, 88*, 291-326.

Proctor, R.W., & Rao, V. (1983a). Null effects of exposure duration and heterogeneity of difference on the 'same-different' disparity in letter matching. *Perception & Psychophysics, 33*, 163-171.

Proctor, R.W., & Rao, V. (1983b). Reinstating the original principles of Proctor's unified theory for matching task phenomena: An evaluation of Krueger and Shapiro's reformulation. *Psychological Review, 90*, 21-37.

Raab, D.H. (1962). Statistical facilitation of simple reaction time. *Transactions of the New York Academy of Sciences, 24*, 574-590.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86*, 446-461.

Ross, S. (1988). *A first course in probability.* New York: Macmillan Publishing Company.

Rubinstein, H., Lewis, S.S., & Rubinstein, M.A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior, 10*, 645-658.

Rumelhart, D.E. & McClelland, J.L. (1982). An interactive activation model of context effects in letter perception: Part 2. *Psychological Review*, 89, 60-94.

Sanders, A.F. (1980). Stage analysis of reaction processes. In: G. Stelmach & J. Requin (Eds.), *Tutorials on motor behavior*. Amsterdam: North Holland.

Sanders, A.F. (1983). Towards a model of stress and human performance. *Acta Psychologica, 53*, 61-97.

Sanders, A.F. (1988). Some issues and trends in the debate on discrete vs. continuous processing of information. Paper presented at a conference concerning Human Information Processing, Amsterdam, september, 1988.

Schmidt, R.A. (1988). *Motor control and learning: A behavioral emphasis*. Champaign (Ill.): Human Kinetics Publishers Inc..

Scinto, L.F.M. (1986). *Written language and psychological development*. Orlando: Academic Press.

Seidenberg, M.S. (1985a). Explanatory adequacy and models of word recognition. *The Behavioral and Brain Sciences, 8*, 724-726.

Seidenberg, M.S. (1985b). The time course of information activation and utilisation in visual word recognition. In D. Besner, T.G. Waller, & G.E. Mackinnon (Eds.), *Reading Research: advances in theory and practice* (Vol. 5), New York: Academic Press, 199-249.

Seidenberg, M.S. (1985c). The time course of phonological code activation in two writing systems. *Cognition, 19*, 1-30.

Seidenberg, M.S. (1987). Sublexical structures in visual word recognition: Access units or orthographic redundancy? In M. Coltheart (Ed.), *Attention and Performance XII*, Hillsdale (N.J.): Lawrence Erlbaum Associates, 245-263.

Seidenberg, M.S., & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.

Seidenberg, M.S., & Tanenhaus, M.K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 546-554.

Seidenberg, M.S., Waters, G.S., Barnes, M.A., & Tanenhaus, M.K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23*, 383-404.

Seif, T.F., & Howard Jr., J.H.(1975). Stimulus compatibility effects of an accessory visual stimulus on auditory sensitivity. *Perception and Psychophysics, 17*, 504-510.

Sejnowski T.J., & Rosenberg, C.R. (1986). NETtalk: A parallel network that learns to read aloud. The John Hopkins University Electrical Engineering and Computer Science Technical Report. JHU/EECS-86/01.

Shallice, T. (1984). More functionally isolable subsystems but fewer "modules"? *Cognition, 17*, 243-252.

Stanislaw, H. (1988). Methodological considerations for the study of multimodal signal detection. *Perception and Psychophysics, 44*, 541-550.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W.G. Koster (Ed.), Attention and Performance II, *Acta Psychologica, 30*, 276-315.

Stroop, J.R. (1935) Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Taft, M. (1986). Lexical access codes in visual and auditory word recognition. *Language and Cognitive Processes, 1*, 297-308.

Tanenhaus, M.K., Flanigan, H.P., & Seidenberg, M.S. (1980). Orthographic and phonological priming in auditory and visual word recognition. *Memory & Cognition, 8*, 513-520.

Thomas, E.A.C. & Ross, B.H. (1980). On appropriate procedures for combining probability distributions within the same family. *Journal of Mathematical Psychology, 21*, 136-152.

Thorson, G., Hochhaus, L., & Stanners, R.F. (1976). Temporal changes in visual and acoustic codes in a letter-matching task. *Perception and Psychophysics, 19*, 346-348.

Titchener, E.B. (1908). *Lectures on the elementary psychology of feeling and attention.* New York: Macmillan.

Tomiak, G.R., Mulennix, J.W., & Sawusch, J.R. (1987). Integral processing of phonemes: Evidence for a phonetic mode of perception. *Journal of the Acoustic Society of America, 81*, 755-764.

Townsend, J.T. & Ashby, F.G. (1983). *The stochastic modeling of elementary psychological processes.* Cambridge: Cambridge University Press.

Treiman, R., Freyd, J., & Baron, J. (1983). Phonological recoding and use of spelling-sound rules in reading of sentences. *Journal of Verbal Learning and Verbal Behavior, 22*, 682-700.

Ulrich, R., & Giray, M. (1986). Separate-activation models with variable base times: Testability and checking of cross-channel dependency. *Perception and Psychophysics, 39*, 248-254.

Van der Heijden, A.H.C. (1981). *Short-term Visual Information Forgetting.* London: Routledge and Kegan Paul.

Van der Heijden, A.H.C., Schreuder, R., Maris, L., & Neerincx, M. (1984). Some evidence for positively correlated separate activation in a simple letter-detection task. *Perception and Psychophysics, 36*, 577-585.

Van Orden, G.C. (1987). A ROWS is a ROSE: Spelling, sound and reading. *Memory & Cognition, 15*, 181-198.

Van Orden, G.C., Johnston, J.C., & Hale, B.L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 371-386.

Venezky, R.L. (1970). *The structure of English orthography.* The Hague: Mouton Press.

Vincent, S.B. (1912). The function of the viborissae in the behavior of the white rat. *Behavioral Monographs, 1*, no. 5.

Vorberg, D. (1989). Workshop "Reaction time models and methods", given at the

Max-Planck-Institute for Psycholinguistics, Nijmegen. February, 6-10.

Vorberg, D. (in preparation). Interacting parallel channels: Predictions of SOA-functions.

Whalen, D.H. (1984). Subcategorical phonetic mismatches slow phonetic judgements. *Perception and Psychophysics, 35*, 49-64.

Wickens, T.D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: Freedman.

Wood, C.C. & Day, R.S. (1975). Failure of selective attention to phonetic segments in consonant-vowel syllables. *Perception and Psychophysics, 17*, 346-350.

Wood, L.E. (1974). Visual and auditory coding in a memory matching task. *Journal of Experimental Psychology, 102*, 106-113.

Wood, L.E. (1977). Modality-specific coding in matching letters. *Perception and Psychophysics, 21*, 329-335.

Yeh, Y., & Eriksen, C.W. (1984). Name codes and features in the discrimination of letter forms. *Perception and Psychophysics, 36*, 225-233.

Zwitserlood, C.M.E. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32*, 25-64.

Curriculum Vitae

Ton Dijkstra werd in 1955 te Tilburg geboren. In 1973 behaalde hij aldaar het diploma Gymnasium β en begon hij zijn studie psychologie aan de Katholieke Hogeschool Tilburg. Na het behalen van het kandidaatsexamen zette hij zijn studie voort aan de Katholieke Universiteit Nijmegen, waar hij in 1980 cum laude afstudeerde in de Psychologische Funktieleer. Vervolgens werkte hij als gewetensbezwaarde bij de vakgroep Interdisciplinaire Onderwijskunde mee aan project SVO 0523, dat zich bezig hield met verschillen in leesvaardigheid tussen goed- en zwak-lezende leerlingen van de lagere school. In 1983 werd hij part-time wetenschappelijk assistent van Dr. U.H. Frauenfelder (Max-Planck-Institut für Psycholinguistik), en part-time onderwijsmedewerker van Prof. Dr. G.A.M. Kempen (Psychologische Funktieleer). In 1986 kende de Max-Planck-Gesellschaft zur Förderung der Wissenschaften (München) hem een tweejarig stipendium toe, waarmee de basis voor deze dissertatie werd gelegd. In 1988 werd zijn aanstelling aan de universiteit omgezet in die van Universitair Docent met volledige dagtaak. Met experimentele ondersteuning vanuit het Max-Planck-Institut werd het onderzoek voor dit proefschrift verder voortgezet en afgerond.

Stellingen

horend bij het proefschrift van Ton Dijkstra
"Cross-modal contacts between graphemes and phonemes".

1. Verwerking van letters of spraakklanken leidt to snelle en automatische activatie van zowel grafeem- als foneemrepresentaties.
[dit proefschrift]

2. Er bestaat geen overtuigend bewijsmateriaal ten gunste van intermodale inhibitie-effecten tussen grafeem- en foneemrepresentaties.
[dit proefschrift]

3. Bimodale sublexicale verwerking kan goed worden beschreven als een gemodificeerde race tussen een visuele en een auditieve stimulus om herkend te worden.
[dit proefschrift]

4. Terwijl complexe en dure onderzoeksmethoden, zoals positron emissie tomografie, een belofte inhouden voor de studie van taalverwerking, zijn ze nog niet gevoelig genoeg om sublexicale intermodale interacties aan te tonen, iets waartoe de relatief eenvoudige en goedkope technieken die nu voorhanden zijn wel in staat zijn.
[dit proefschrift; Peterson, S.E., Fox, P.T., Posner, M.I., Mintun, M., & Raichle, M.E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331*, 585-589]

5. Hoewel computersimulatie de specificatie vereist van talrijke niet altijd empirisch gemotiveerde aannames, is het een onmisbaar stuk gereedschap voor de formulering en toetsing van complexere theorieën op het gebied van de psycholinguïstiek.
[dit proefschrift]

6. Vanuit psychologisch en maatschappelijk standpunt is het wenselijk op de basis- en middelbare school een vak als "emotionele en sociale opvoeding" in te voeren.

7. Life is not a problem to be solved, but a mystery to be lived.
[Osho]

8. De afgelopen drie maanden hebben geleerd dat roken niet slechts lijkt te kunnen leiden tot de reeds bekende negatieve gevolgen voor de gezondheid, maar tevens tot selectieve leesblindheid, gehoorstoornissen en geheugenverlies.